

POLSKA AKADEMIA NAUK
INSTYTUT PODSTAWOWYCH PROBLEMÓW TECHNIKI

ARCHIWUM ELEKTROTECHNIKI

TOM V · ZESZYT 2

KWARTALNIK

PAŃSTWOWE WYDAWNICTWO NAUKOWE
WARSZAWA 1956

R A D A R E D A K C Y J N A

PROF. JANUSZ LECH JAKUBOWSKI, PROF. BOLESŁAW KONORSKI,
PROF. IGNACY MALECKI, PROF. WITOLD NOWICKI, PROF. STANISŁAW SZPOR
PROF. PAWEŁ SZULKIN

K O M I T E T R E D A K C Y J N Y

Redaktor Naczelny

PROF. JANUSZ GROSZKOWSKI

Z-ca Redaktora Naczelnego

PROF. PAWEŁ NOWACKI

Sekretarz

WACŁAW ZWOLAKOWSKI

ADRES REDAKCJI:

Warszawa, ul. Koszykowa 75, Politechnika,
Zakład Elektroniki PAN, tel. 8.32.04

Państwowe Wydawnictwo Naukowe - Dział Czasopism
Warszawa, Krakowskie Przedmieście 79

Nakład 1490+104 egz.	Do składania 13 III 1956
Ark. wyd. 16,65 druk. 13,75	Podpisano do druku 8 VI 1956
Papier dr. sat. 70 g V kl. 70×100/16	Druk ukończono w czerwcu 1956
Cena zł 20.—	K-7-2186 - Zamówienie 109/165

POZNAŃSKA DRUKARNIA NAUKOWA — POZNAŃ, FREDRY 10

621.3.011.4:621.3.015

B. KONORSKI

Kąty graniczne w polu elektrostatycznym dwóch kul o ładunkach rozmaitych znaków

Rękopis dostarczono 27. 6. 1955

Przedmiotem rozważań jest pole elektrostatyczne układu dwóch kul, na których znajdują się dwa ładunki o rozmaitych znakach i o rozmaitych wartościach bezwzględnych. Wszystkie linie pola wychodzące z ładunku o mniejszej wartości bezwzględnej kończą się na powierzchni drugiej kuli wyznaczając na niej okrąg położony symetrycznie względem osi łączącej środki obu kul. Proste poprowadzone z punktów tego okręgu do środka kuli tworzą stożek kołowy o kącie wierzchołkowym 2ϑ . Kąt ϑ nazwany jest kątem granicznym; dzieli on powierzchnię kuli na dwie części; linie pola wychodzące z punktów jednej części dążą do drugiej kuli, linie pola wychodzące z punktów drugiej części dążą do nieskończoności.

W pracy podana jest metoda, za pomocą której można obliczać kąty graniczne oraz przeprowadzone są obliczenia tych kątów przy rozmaitych stosunkach bezwzględnych wartości ładunków lub potencjałów obu kul.

1. STRUMIEŃ ELEKTRYCZNY ŁADUNKU PRZENIKAJĄCY PRZEZ CZASZĘ KULISTĄ

Rozważmy najpierw następujące proste zadanie. Należy obliczyć strumień ψ punktowego ładunku elektrycznego Q przenikający czaszę kulistą CEF o promieniu r i o kącie centralnym 2λ . Przyjmując oznaczenia z rys. 1, w którym założono $O_1O = A$; $OP = r_0$; $O_1P = \varrho$; $O_1B \perp OP$; $BP = a$, znajdujemy przesunięcie D w punkcie P

$$-D = -\frac{Q}{4\pi\varrho^2} \cos \theta = -\frac{Q}{4\pi\varrho^2} \cdot \frac{a}{\varrho} = \frac{-Q(A \cos \vartheta - r_0)}{4\pi\varrho^3}.$$

Oznaczając przez $d\psi_z$ strumień przenikający przez pierścień czaszy między kątami ϑ i $\vartheta + d\vartheta$ mamy

$$d\psi_z = -\frac{Q(A \cos \vartheta - r_0)}{4\pi\varrho^3} 2\pi r_0 \sin \vartheta \cdot r_0 d\vartheta = -\frac{Qr_0^2}{2} \cdot \frac{A \cos \vartheta - r_0}{\varrho^3} \sin \vartheta d\vartheta$$

$$\psi_z = -\frac{Qr_0^2}{2} \int_0^\lambda \frac{(A \cos \vartheta - r_0) \sin \vartheta d\vartheta}{\sqrt{r_0^2 + A^2 - 2r_0A \cos \vartheta^3}}$$

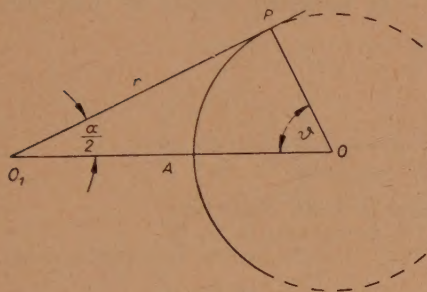
i po przeliczeniach otrzymujemy

$$\psi_w = + \int_{2\lambda} \bar{D} \cdot d\bar{S} = \frac{Q}{2} + \frac{Q}{4A} \cdot \frac{\varrho_\lambda^2 - r_0^2 + A^2}{\varrho_\lambda}. \quad (2)$$

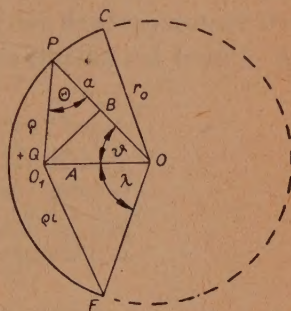
Ze wzoru (2)

$$\text{dla } \lambda=0 \quad \text{i} \quad \varrho_\lambda = r_0 - A \quad \text{jest} \quad \psi_w = 0,$$

$$\text{dla } \lambda=\pi \quad \text{i} \quad \varrho_\lambda = r_0 + A \quad \text{jest} \quad \psi_w = \psi_{w \max} = Q.$$



Rys. 2. Maksymalny strumień dielektryczny przenikający przez czaszę kulistą.



Rys. 3. Przenikający przez czaszę kulistą strumień dielektryczny ładunku położonego wewnątrz kuli.

Równania (1) i (2) możemy także przedstawić w postaci

$$\psi_z = \frac{Q}{2} \left(\frac{A - r_0 \cos \vartheta}{\varrho} - 1 \right), \quad (3)$$

$$\psi_w = \frac{Q}{2} \left(\frac{A - r_0 \cos \vartheta}{\varrho} + 1 \right) \quad (4)$$

(zamiast λ piszemy tu ϑ).

Jeżeli ładunek Q umieszczony jest w środku czaszy (w punkcie O), to strumień ψ_w wynosi

$$\psi_w = -\frac{Q}{2} (\cos \vartheta - 1). \quad (5)$$

2. PRZYBLIŻONE ROZMIESZCZENIE ZASTĘPCZE ŁADUNKÓW W UKŁADZIE DWÓCH KUL O POTENCJAŁACH V , O

Założmy podobnie jak w [A, rozdz. 7]¹, że mamy układ dwóch kul metalowych K , K' ; odległość środków tych kul wynosi D . Na kuli K

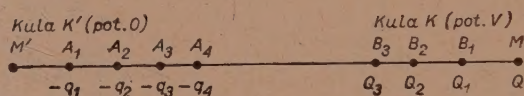
¹ Przez [A] oznaczamy tu artykuł autora pt.: „Nowe twierdzenia o polu elektrostatycznym” (Arch. Elektr. t. IV, z. 1); liczby umieszczone bezpośrednio przy A oznaczają numer wzoru, na który się powołujemy.

(promień r) znajduje się ładunek $+Q$ (który nazywamy ładunkiem s w o b o d n y m) sprawiający, że potencjał tej kuli jest

$$V = \frac{Q}{4\pi\epsilon r};$$

na kuli K' (promień r') swobodnego ładunku nie ma, jej potencjał V równa się zeru.

Wiemy, że układ powyższy można zastąpić układem ładunków punktowych umieszczonych na prostej $M'-M$ łączącej środki M' i M obu kul K' i K . Układ ten składa się (rys. 4)



Rys. 5. Zastępcze rozmieszczenie ładunków w układzie dwóch kul o potencjałach V , O .

z ładunków dodatnich $Q, Q_1, Q_2, Q_3 \dots$

umieszczonych w punktach $M, B_1, B_2, B_3 \dots$

oraz

z ładunków ujemnych $-q_1, -q_2, -q_3, -q_4 \dots$

umieszczonych w punktach $A_1, A_2, A_3, A_4 \dots$

Wiadomo następnie, że między odległościami β_n punktów B_n od M i odległościami a_n punktów A_n od M' zachodzą związki [A 28]

$$a_n = \frac{r'^2}{D - \beta_{n-1}}; \quad \beta_n = \frac{r^2}{D - a_n}; \quad (n=1, 2, 3, \dots; \beta_0=0); \quad (6)$$

między ładunkami Q_n i q_n zachodzą związki [A 28]

$$Q_n = q_n \frac{r}{D - a_n}; \quad q_n = Q_{n-1} \frac{r'}{D - \beta_{n-1}} \quad (n=1, 2, 3, \dots; Q_0=Q). \quad (7)$$

Związki powyższe prowadzą do nierówności

$$a_1 < a_2 < a_3 \dots,$$

$$\beta_1 < \beta_2 < \beta_3 \dots,$$

$$Q > Q_1 > Q_2 > Q_3 \dots,$$

$$q_1 > q_2 > q_3 \dots$$

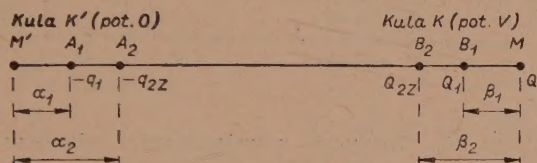
Jak wynika z odnośnych obliczeń, wyrazy ostatnich dwóch ciągów szybko maleją. Szybko maleją również różnice $(a_n - a_{n-1})$ oraz $(\beta_n - \beta_{n-1})$ przy

wzrastaniu n . Na podstawie tych własności pierwsze dwa ciągi zastępujemy ciągami

$$\alpha_1, \alpha_2, \alpha_2, \alpha_2 \dots$$

$$\beta_1, \beta_2, \beta_2, \beta_2 \dots$$

i zamiast dokładnego rozkładu ładunków przedstawionego na rys. 4 otrzymujemy przybliżony rozkład zastępczy ładunków umieszczonych w punktach M, B_1, B_2, A_1, A_2 (rys. 5), przy czym (patrz [A 54] i [A 57]):



Rys. 5. Zastępcze rozmieszczenie ładunków w układzie dwóch kul o potencjałach $V, 0$.

w punkcie M umieszczony jest ładunek Q ,

$$,, \quad ,, \quad B_1 (B_1 M = \beta_1) \quad ,, \quad ,, \quad Q_1 = Qp, \quad (7)$$

$$,, \quad ,, \quad B_2 (B_2 M = \beta_2) \quad ,, \quad ,, \quad Q_{2z} = SQ - Q - Q_1, \quad (8)$$

$$Q_{2z} = SQ - Q - Q_1 = Q(S - 1 - p) = Qn,$$

$$\text{w punkcie } A_1 (A_1 M = \alpha_1) \text{ umieszczony jest ładunek } -q_1 = -Q \frac{r'}{D}, \quad (9)$$

$$,, \quad ,, \quad A_2 (A_2 M = \alpha_2) \quad ,, \quad ,, \quad -q_{2z},$$

$$q_{2z} = QT - q_1 = Q \left(T - \frac{r'}{D} \right) = Q \frac{r'}{2kD} (1 + R) = \frac{Qmr'}{D}. \quad (10)$$

Analogicznie do [A 39, 44, 77, 79 i 89] wprowadzono tu następujące skróty:

$$p = \frac{rr'}{D^2 - r'^2}, \quad (11)$$

$$R = \sqrt{\frac{D^2 - (r' - r)^2}{D^2 - (r' + r)^2}}, \quad k = \frac{D^2 - r^2 - r'^2}{rr'} \quad (12)$$

oraz

$$n = \frac{p}{2} (R - 1), \quad m = \frac{1}{2k} (R + 1), \quad (13)$$

$$S = 1 + p + n, \quad T = \frac{r'}{D} (m + 1). \quad (14)$$

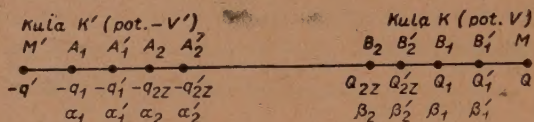
Wobec tego mamy

$$\alpha_1 = \frac{r'^2}{D}, \quad \alpha_2 = \frac{r'^2}{Dkp}, \quad (15)$$

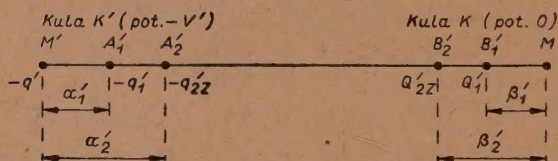
$$\beta_1 = \frac{Drp}{r'}, \quad \beta_2 = \frac{r^2 Dkp}{D^2 kp - r'^2}. \quad (16)$$

3. PRZYBLIŻONE ROZMIESZCZENIE ZASTĘPCZE ŁADUNKÓW W UKŁADZIE DWÓCH KUL O POTENCJAŁACH $-V'$, $+V$

Jeżeli obie kule mają potencjały $-V'$ i $+V$, to odpowiednie rozmieszczenie ładunków znajdujących się na prostej MM' otrzymuje się przez superpozycję układu ładunków przy potencjałach O i V oraz układu ła-



Rys. 6. Zastępcze rozmieszczenie ładunków w układzie dwóch kul o potencjałach O , $-V'$.



Rys. 7. Zastępcze rozmieszczenie ładunków w układzie dwóch kul o potencjałach V , $-V'$.

dunków przy potencjałach $-V'$ i O . Również i zastępcze rozmieszczenie przybliżone ładunków otrzymamy przez superpozycję ładunków z rys. 5 i 6; rozmieszczenie to przedstawia rys. 7. Mamy tu

$$\alpha_1 = \frac{r'^2}{D}, \quad \beta_1 = \frac{r^2}{D - \alpha_1}, \quad \alpha_2 = \frac{r'^2}{D - \beta_1}, \quad \beta_2 = \frac{r^2}{D - \alpha_2}, \quad (17)$$

$$\beta_1' = \frac{r^2}{D}, \quad \alpha_1' = \frac{r'^2}{D - \beta_1'}, \quad \beta_2' = \frac{r^2}{D - \alpha_1'}, \quad \alpha_2' = \frac{r'^2}{D - \beta_2'}. \quad (18)$$

Wynika stąd

$$\begin{aligned} \alpha_1' &> \alpha_1, & \beta_1 &> \beta_1', \\ D - \alpha_1 &> D - \alpha_1', & D - \beta_1 &< D - \beta_1', \\ \beta_1 &< \beta_2', & \alpha_2 &> \alpha_1', \\ D - \beta_1 &> D - \beta_2', & D - \alpha_2 &< D - \alpha_1', \\ \alpha_2 &< \alpha_2', & \beta_2 &> \beta_2', \\ \alpha_1 &< \alpha_1' < \alpha_2 < \alpha_2', & \beta_1' &< \beta_1' < \beta_2' < \beta_2'. \end{aligned}$$

skąd

Wprowadzamy oznaczenia

$$p' = \frac{rr'}{D^2 - r^2}, \quad n' = \frac{p'}{2}(R-1), \quad (19)$$

$$S' = 1 + p' + n', \quad T' = \frac{r}{D}(m+1) \quad (20)$$

i znajdujemy wartości ładunków umieszczonych w poszczególnych punktach schematu z rys. 7.

Kula K'			Kula K		
Punkt	M' ładunek	$-q'$	Punkt	M ładunek	Q
A_1	„	$-Q \frac{r'}{D} = -q_1$,	„	B'_1	„ $q' \frac{r}{D} = Q'_1$,
A'_1	„	$-q'p' = -q'_1$,	„	B_1	„ $Qp = Q_1$,
A_2	„	$-Q \frac{r'm}{D} = -q_{2z}$,	„	B'_2	„ $q' \frac{rm}{D} = Q'_{2z}$
A'_2	„	$-q'n' = -q'_{2z}$,	„	B_2	„ $Qn = Q_{2z}$.

Łączny ładunek H wewnątrz kuli K wynosi (por. [A 96])

$$H = Q + Qp + Qn + q' \frac{r}{D}(m+1) = QS + q'T', \quad (21)$$

a łączny ładunek H' wewnątrz kuli K wynosi (por. [A 95])

$$H' = q' + q'p' + q'n' + Q \frac{r'}{D}(m+1) = q'S' + QT'. \quad (22)$$

Jeżeli promienie r, r' obu kuli są jednakowe, to

$$a_1 = \frac{r^2}{D} = \beta'_1, \quad a'_1 = \frac{r^2}{D - a_1} = \beta_1, \quad (23)$$

$$\beta'_2 = \frac{r^2}{D - \beta_1} = a_2, \quad a'_2 = \frac{r^2}{D - a^2} = \beta_2. \quad (24)$$

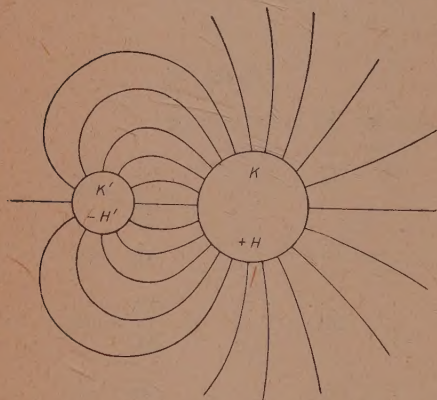
4. KĄT GRANICZNY W UKŁADZIE DWÓCH KUL O POTENCJAŁACH V i $-V'$

Założmy, że absolutna wartość ładunku kuli K' jest mniejsza niż wartość ładunku kuli K , to znaczy, że

$$|H'| < |H|. \quad (25)$$

Z rozważań wynika, że przy tym założeniu wszystkie linie pola wychodzące z powierzchni kuli K' kończą się na powierzchni kuli K , jednakże pewna ilość linii pola wychodzących z powierzchni kuli K dąży do nieskończoności (rys. 8); linie te łącznie odpowiadają ładunkowi $(H - H')$.

Założmy, że na obu kulach K i K' (o wymiarach skończonych) znajdują się ładunki $+H'$ i $-H'$ o równych wartościach bezwzględnych i od-



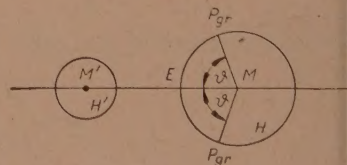
Rys. 8. Schematyczny obraz pola elektrostatycznego między dwiema kulami o ładunkach $+H$ i $-H'$. ($H > H'$).

miennych znakach. W bardzo odległych punktach P przestrzeni, w której znajduje się układ K, K' , natężenia pól wytworzonych przez oba ładunki $+H$ i $-H'$ mają w przybliżeniu przeciwne zwroty i jednakowe wartości² (dążące do zera w miarę oddalania się punktów P od układu K, K'). Wynika stąd, że natężenie pola układu K, K' równa się zero na nieskończenie odległej sferze. Również i składowa natężenia tego pola prostopadła do tej sfery równa się zero, co oznacza, że żadna linia pola wychodząca z elementów układu K, K' nie kończy się w nieskończoności, oraz dalej, że wszystkie linie pola znajdujące swe ujście na K' wychodzą z K , jak

również, że wszystkie linie pola wychodzące z K kończą się na K .

Jeśli teraz, pozostawiając bez zmiany ładunek kuli K' , dodamy do kuli K dodatkowy dodatni ładunek $(H - H')$, to linie pola tego dodatkowego ładunku nie mogą kończyć się na kuli K' (ponieważ oznaczałoby to zwiększenie ujemnego ładunku tej kuli) i muszą dążyć do nieskończoności. Natomiast bez zmiany pozostaje stwierdzony już poprzednio fakt, że wszystkie linie kończące się na K' wychodzą z K .

Pole elektrostatyczne z rys. 8 ma symetrię osiową; osią jest tu prosta $M'M$. Wynika stąd, że na kuli musi istnieć taki okrąg



Rys. 9. Kąt graniczny w polu dwóch kul o niejednakowych ładunkach $+H$ i $-H'$.

² Przekonujemy się o tym zastępując pole ładunków $+H'$, $-H'$ na kulach K, K' przez dwa ciągi dipoli:

$$+Q_1 \text{ i } q'_1, +Q_1 \text{ i } -q'_1, +Q_2 \text{ i } -q'_2 \dots \dots \text{ itd.}$$

oraz

$$+Q'_1 \text{ i } -q_1, +Q'_2 \text{ i } -q_2 \text{ itd.}$$

(wartości bezwzględne każdej pary tych ładunków są równe).

$P_{gr} - P_{gr}$ (o środku na prostej $M'M$, rys. 9), mający tę własność, iż wszystkie linie pola wychodzące z czasy kulistej $P_{gr} EP_{gr}$ kończą się na kuli K' , podczas gdy wszystkie linie wychodzące z pozostałej części kuli K dążą do nieskończoności. Łącząc środek M z punktami okręgu $P_{gr} - P_{gr}$ otrzymujemy stożek kołowy o kącie wierzchołkowym 2ϑ . Kąt ϑ nazywamy kątem granicznym w rozważanym polu dwóch kul.

Strumień ψ_ϑ przenikający przez czasę kulistą $P_{gr}EP_{gr}$ można w przybliżeniu obliczyć, jeżeli układ z rys. 8 zastąpić układem ładunków punktowych rozmieszczonych według rys. 7. Stosując do ładunków $-q'$, $-q$ wzory (3), do ładunków Q_i wzory (4) i do ładunku Q wzór (5) znajdujemy

$$\begin{aligned} \psi_\vartheta = & -\frac{q'}{2} \left(\frac{D-r \cos \vartheta}{\varrho'_M} - 1 \right) - \frac{Qr'}{2D} \left(\frac{D-a_1-r \cos \vartheta}{\varrho_{a1}} - 1 \right) - \\ & - \frac{q'p'}{2} \left(\frac{D-a'_1-r \cos \vartheta}{\varrho'_{a1}} - 1 \right) - Q \frac{r'(R+1)}{4kD} \left(\frac{D-a_2-r \cos \vartheta}{\varrho_{a2}} - 1 \right) \\ & - \frac{q'p'(R-1)}{4} \left(\frac{D-a'_2-r \cos \vartheta}{\varrho'_{a2}} - 1 \right) + pQ \frac{R-1}{4} \left(\frac{\beta_2-r \cos \vartheta}{\varrho_{\beta 2}} + 1 \right) \quad (26) \\ & + q' \frac{rm}{2D} \left(\frac{\beta'_2-r \cos \vartheta}{\varrho'_{\beta 2}} + 1 \right) + \frac{Qp}{2} \left(\frac{\beta_1-r \cos \vartheta}{\varrho_{\beta 1}} + 1 \right) \\ & + q' \frac{\tau}{2D} \left(\frac{\beta'_1-r \cos \vartheta}{\varrho'_{\beta 1}} + 1 \right) + \frac{Q}{2} (-\cos \vartheta). \end{aligned}$$

W równaniu powyższym ϱ'_M jest odległością punktu M (rys. 7) od punktu P_{gr} (rys. 9), $\varrho_{a1}, \varrho'_{a1}, \varrho_{a2}, \varrho'_{a2}$ — odległościami punktów A_1, A'_1, A_2, A'_2 od punktu P_{gr} , a $\varrho_{\beta 2}, \varrho'_{\beta 2}, \varrho_{\beta 1}, \varrho'_{\beta 1}$ — odległościami punktów B_2, B'_2, B_1, B'_1 od punktu P_{gr} . Wyrażenie (26) można uprościć uwzględniając, że potencjał punktu P_{gr} jest V ; możemy zatem napisać tożsamość

$$\begin{aligned} \frac{1}{4\pi\epsilon_0} \left(-\frac{q'}{\varrho'_M} - \frac{Qr'}{D\varrho_{a1}} - \frac{q'p'}{\varrho'_{a1}} - \frac{Qr'(R+1)}{2kD\varrho_{a2}} - \frac{q'p'(R-1)}{2\varrho'_{a2}} + \frac{pQ(R-1)}{2\varrho_{\beta 2}} + \right. \\ \left. + \frac{q'rm}{D\varrho_{\beta 2}} + \frac{Qp}{\varrho_{\beta 1}} + \frac{q'r}{D\varrho_{\beta 1}} + \frac{Q}{r} \right) = V = \frac{Q}{4\pi\epsilon_0 r}. \quad (27) \end{aligned}$$

Tożsamość (27) musi być spełniona niezależnie od wartości ładunków Q i q . Wynika stąd, że współczynniki przy tych wielkościach muszą być równe zeru:

$$-\frac{1}{\varrho'_M} - \frac{p'}{\varrho'_{a1}} - \frac{n'}{\varrho'_{a2}} + \frac{rm}{D\varrho'_{\beta 2}} + \frac{r'}{D\varrho'_{\beta 1}} = 0, \quad (28)$$

$$-\frac{r'}{D\varrho_{a1}} - \frac{r'm}{D\varrho_{a2}} + \frac{n}{\varrho_{\beta 2}} + \frac{p}{\varrho_{\beta 1}} = 0. \quad (29)$$

Korzystając ze związków (28) i (29) wyrażenie ψ_θ możemy przedstawić w postaci

$$\begin{aligned} \psi_\theta = & -\frac{q'}{2} \left(\frac{D}{\varrho'_M} - 1 \right) - \frac{Qr'}{2D} \left(\frac{D-a_1}{\varrho_{a1}} - 1 \right) - \frac{q'p'}{2} \left(\frac{D-a'_1}{\varrho'_{a1}} - 1 \right) - \\ & - \frac{Qr'm}{2D} \left(\frac{D-a_2}{\varrho_{a2}} - 1 \right) - \frac{q'n'}{2} \left(\frac{D-a'_2}{\varrho'_{a2}} - 1 \right) + \frac{Qn}{2} \left(\frac{\beta_2}{\varrho_{\beta 2}} + 1 \right) + \\ & + \frac{q'rm}{2D} \left(\frac{\beta'_2}{\varrho'_{\beta 2}} + 1 \right) + \frac{Qp}{2} \left(\frac{\beta_1}{\varrho_{\beta 1}} + 1 \right) + \frac{q'r}{2D} \left(\frac{\beta'_1}{\varrho'_{\beta 1}} + 1 \right) + \frac{Q}{2} (1 - \cos \vartheta), \end{aligned}$$

lub po dokonaniu drobnych przekształceń

$$\begin{aligned} \psi_\theta = & -\frac{q'}{2} \frac{D}{\varrho'_M} - \frac{Qr'}{2D} \frac{D-a_1}{\varrho_{a1}} - \frac{q'p'}{2} \frac{D-a'_1}{\varrho'_{a1}} - \frac{Qr'm}{2D} \frac{D-a_2}{\varrho_{a2}} - \\ & - \frac{q'n'}{2} \frac{D-a'_2}{\varrho'_{a2}} + \frac{Qn}{2} \frac{\beta_2}{\varrho_{\beta 2}} + \frac{q'rm}{2D} \frac{\beta'_2}{\varrho'_{\beta 2}} + \frac{Qp}{2} \frac{\beta_1}{\varrho_{\beta 1}} + \frac{q'r\beta'_1}{2D\varrho'_{\beta 1}} + \\ & + \frac{q'}{2} (1+p'+n') + \frac{Q}{2} (1+p+n) + \frac{Qr'+q'r}{2D} (m+1) - \frac{Q}{2} \cos \vartheta. \quad (30) \end{aligned}$$

Z poczynionych założeń wynika, że strumień ψ_θ musi się równać całkowitemu ładunkowi H' kuli K' , a przeto możemy napisać

$$\psi_\theta = S'q' + TQ = q'(1+p'+n') + Q \frac{r'}{D} (m+1). \quad (31)$$

Podstawiając ψ_θ ze wzoru (30) znajdujemy

$$\begin{aligned} & -\frac{q'}{2} \frac{D}{\varrho'_M} - \frac{Qr'}{2D} \frac{D-a_1}{\varrho_{a1}} - \frac{q'p'}{2} \frac{D-a'_1}{\varrho'_{a1}} - \frac{Qr'm}{2D} \frac{D-a_2}{\varrho_{a2}} - \frac{q'n'}{2} \frac{D-a'_2}{\varrho'_{a2}} + \\ & + \frac{Qn}{2} \frac{\beta_2}{\varrho_{\beta 2}} + \frac{q'rm}{2D} \frac{\beta'_2}{\varrho'_{\beta 2}} + \frac{Qp}{2} \frac{\beta_1}{\varrho_{\beta 1}} + \frac{q'r}{2D} \frac{\beta'_1}{\varrho'_{\beta 1}} + \\ & + \frac{Q}{2} (1+p+n) - \frac{q'}{2} (1+p'+n') + \frac{q'r - Qr'}{2D} (m+1) - \frac{Q}{2} \cos \vartheta = 0. \quad (32) \end{aligned}$$

Przy danych parametrach geometrycznych układu obu kul oraz ładunkach Q , q' , z równania (32) można obliczyć kąt graniczny ϑ .

5. OBLICZENIE KĄTA GRANICZNEGO

W celu obliczenia kąta ϑ z równania (32) rozwijamy wchodzące w to wyrażenie odwrotności ϱ za pomocą wielomianów Legendre'a $P_0, P_1, P_2, \dots$. Piszemy na przykład

$$\begin{aligned} \frac{1}{\varrho_{a1}} &= \frac{1}{\sqrt{(D-a_1)^2 - 2(D-a_1)r \cos \vartheta + r^2}} = \frac{1}{D-a_1} \frac{1}{\sqrt{1 - 2 \frac{r}{D-a_1} \cos \vartheta + \left(\frac{r}{D-a_1}\right)^2}} \\ &= \frac{1}{D-a_1} \left[P_0 + \frac{r}{D-a_1} P_1 + \frac{r^2}{(D-a_1)^2} P_2 + \dots \right], \end{aligned} \quad (33)$$

$$\begin{aligned} \frac{1}{\varrho_{\beta 1}} &= \frac{1}{\sqrt{r^2 - 2\beta_1 r \cos \vartheta + \beta_1^2}} = \frac{1}{r} \frac{1}{\sqrt{1 - 2 \frac{\beta_1}{r} \cos \vartheta + \left(\frac{\beta_1}{r}\right)^2}} = \\ &= \frac{1}{r} \left[P_0 + \frac{\beta_1}{r} P_1 + \frac{\beta_1^2}{r^2} P_2 + \dots \right]. \end{aligned} \quad (34)$$

Oznaczając

$$\begin{aligned} \frac{D}{\varrho_M} = F_1, \quad \frac{D-a_1}{\varrho_{a1}} = F_2, \quad \frac{D-a_1'}{\varrho_{a1}} = F_3, \quad \frac{D-a_2}{\varrho_{a2}} = F_4, \quad \frac{D-a_2'}{\varrho_{a2}} = F_5, \\ \frac{r}{\beta_2} = F_6, \quad \frac{r}{\beta_2'} = F_7, \quad \frac{r}{\beta_1} = F_8, \quad \frac{r}{\beta_1'} = F_9 \end{aligned} \quad (35)$$

otrzymujemy 5 wyrażeń typu (33) i 4 wyrażenia typu (34). Dzięki zależności

$$\frac{r}{D-a_1} = \frac{\beta_1}{r}, \quad \frac{r}{D-a_1'} = \frac{\beta_2'}{r}, \quad \frac{r}{D-a_2} = \frac{\beta_2}{r}, \quad \frac{\beta_1'}{r} = \frac{r}{D},$$

niektóre z wyrażeń F są sobie równe, a mianowicie:

$$F_2 = F_8, \quad F_4 = F_6, \quad F_9 = F_1, \quad F_7 = F_3. \quad (36)$$

Otrzymujemy zatem 5 szeregów

$$\begin{aligned} F_1 &= 1 + \frac{r}{D} \cos \vartheta + \frac{r^2}{D^2} P_2 + \frac{r^3}{D^3} P_3 + \dots, \\ F_6 &= 1 + \frac{\beta_2}{r} \cos \vartheta + \frac{\beta_2^2}{r^2} P_2 + \frac{\beta_2^3}{r^3} P_3 + \dots, \end{aligned} \quad (37)$$

$$\begin{aligned} F_8 &= 1 + \frac{\beta_1}{r} \cos \vartheta + \frac{\beta_1^2}{r^2} P_2 + \frac{\beta_1^3}{r^3} P_3 + \dots, \\ F_3 &= 1 + \frac{r}{D-a_1'} \cos \vartheta + \left(\frac{r}{D-a_1'}\right)^2 P_2 + \left(\frac{r}{D-a_1'}\right)^3 P_3 + \dots, \end{aligned} \quad (38)$$

$$F_5 = 1 + \frac{r}{D-a_2} \cos \vartheta + \left(\frac{r}{D-a_2}\right)^2 P_2 + \left(\frac{r}{D-a_2}\right)^3 P_3 + \dots \quad (39)$$

Jeżeli $r=r'$, to na skutek

$$\frac{r}{D-a_1} = \frac{r}{D-\beta_1} = \frac{a_2}{r}$$

zamiast (38) możemy napisać

$$F_3 = 1 + \frac{a_2}{r} \cos \vartheta + \frac{a_2^2}{r^2} P_2 + \frac{a_2^3}{r^3} P_3 + \dots \quad (40)$$

Jeżeli zaś $r=r'$ i $\beta_2 \approx a_2$, to w przybliżeniu możemy założyć

$$\frac{r}{D-a_2} = \frac{r}{D-\beta_2} \approx \frac{r}{D-a_2} = \frac{\beta_2}{r}$$

i wtedy

$$F_3 = F_5 = F_6. \quad (41)$$

Przy wykorzystaniu wzorów (35) i (36) równanie (32) możemy przedstawić jako

$$\begin{aligned} \Delta = Q \left[\left(\frac{\beta_2 n}{r} - \frac{r' m}{D} \right) F_6 + \left(\frac{\beta_1 p}{r} - \frac{r'}{D} \right) F_8 + 1 + p + n - \frac{r'}{D} (m+1) - \cos \vartheta \right] + \\ + q' \left[\left(\frac{\beta_2' m}{D} - p' \right) F_3 + \left(\frac{\beta_1'}{D} - 1 \right) F_1 - n' F_5 - 1 - p' - n' + \frac{r}{D} (m+1) \right] = 0, \end{aligned} \quad (42)$$

a wyrażenie (30) dla ψ_ϑ przybiera postać

$$\begin{aligned} \psi_\vartheta = \frac{Q}{2} \left[\left(\frac{\beta_2 n}{r} - \frac{r' m}{D} \right) F_6 + \left(\frac{\beta_1 p}{r} - \frac{r'}{D} \right) F_8 + 1 + p + n + \frac{r'}{D} (m+1) - \cos \vartheta \right] + \\ + \frac{q'}{2} \left[\left(\frac{\beta_2' m}{D} - p' \right) F_3 + \left(\frac{\beta_1'}{D} - 1 \right) F_1 - n' F_5 + 1 + p' + n' + \frac{r}{D} (m+1) \right]. \end{aligned} \quad (43)$$

Równanie (42) służy do bezpośredniego obliczenia kąta granicznego ϑ . Przykład takiego obliczenia przedstawiony jest niżej.

Samo zestawienie równania (42) jest dość skomplikowane. Duża liczba występujących w nim współczynników i parametrów prowadzi do nieuchronnych prawie omyłek rachunkowych. Z tego powodu konieczne jest opracowanie metod, za pomocą których można sprawdzić słuszność zestawionego równania liczbowego. Do tego celu służy równanie (43). W pracy zostały podane dwie metody, przy czym ze względu na kapryśny charakter współczynników liczbowych, zaleca się przy przeprowadzanych rachunkach stosowanie obu sposobów sprawdzania.

Sprawdzenie I

Podstawiając do (43) $\vartheta=0$ powinno się otrzymać wynik

$$\psi_\vartheta = 0. \quad (44)$$

Relacja powyższa prowadzi do tożsamości

$$\left[\left(\frac{\beta_2 n}{r} - \frac{r' m}{D} \right) F_6 + \left(\frac{\beta_1 p}{r} - \frac{r'}{D} \right) F_8 \right]_{\vartheta=0} + S + T - 1 = 0, \quad (45)$$

$$\left[\left(\frac{\beta'_2 m}{D} - p' \right) F_3 + \left(\frac{\beta'_1}{D} - 1 \right) F_1 - n' F_5 \right]_{\vartheta=0} + S' + T' = 0$$

albo przy $r=r'$

$$1 - \left[\left(\frac{\beta_2 n}{r} - \frac{r m}{D} \right) F_6 + \left(\frac{\beta_1 p}{r} - \frac{r}{D} \right) F_8 \right]_{\vartheta=0} = - \left[\left(\frac{\alpha_2 m}{D} - p \right) F_3 + \right. \\ \left. + \left(\frac{a_1}{D} - 1 \right) F_1 - n F_5 \right]_{\vartheta=0} = S + T. \quad (46)$$

Sprawdzenie II

Jeżeli $\vartheta = \pi$, to musi być $H = H'$, a przeto musi być spełniona tożsamość

$$\psi_{\vartheta} = H' = QS + q'T = (1 + p + n)Q + r' \frac{m+1}{D} q'. \quad (47)$$

Z (21) i (22) w tym przypadku otrzymujemy

$$q' = Q \frac{S - T}{S' - T'}.$$

Podstawiając to wyrażenie do (43) znajdujemy

$$\left[\left(\frac{\beta_2 n}{r} - \frac{r' m}{D} \right) F_6 + \left(\frac{\beta_1 p}{r} - \frac{r'}{D} \right) F_8 - \cos \vartheta \right]_{\vartheta=\pi} \cdot (S' - T') + \\ + \left[\left(\frac{\beta'_2 m}{D} - p' \right) F_3 + \left(\frac{\beta'_1}{D} - 1 \right) F_1 - n' F_5 \right]_{\vartheta=\pi} \cdot (S - T) = 0$$

lub przy $r=r'$

$$- \left[\left(\frac{\beta_2 n}{r} - \frac{r m}{D} \right) F_6 + \left(\frac{\beta_1 p}{r} - \frac{r}{D} \right) F_8 - \cos \vartheta \right]_{\vartheta=\pi} = \\ = \left[\left(\frac{\beta_2 m}{D} - p \right) F_3 + \left(\frac{\beta_1}{D} - 1 \right) F_1 - n F_5 \right]_{\vartheta=\pi}. \quad (49)$$

Można wykazać (patrz Dodatek), że przybliżona wartość wyrażeń znajdujących się po obu stronach równości (49) wynosi $(S - T)$.

6. PRZYKŁADY OBLICZEŃ

A. Przykład obliczeń

gdy dane $D=5$, $r=r'=1$.

(a)

1. Zestawienie równania (42)

$$\frac{r}{D} = 0,2, \quad p = \frac{1}{25-1} = 0,04167, \quad k = \frac{25-1-1}{1} = 23,$$

$$R = \frac{5}{\sqrt{25-4}} = 1,09109, \quad R+1 = 2,09109, \quad R-1 = 0,09109,$$

$$m = \frac{2,09109}{2 \cdot 23} = 0,04546, \quad n = \frac{0,09109}{48} = 0,00190,$$

$$\alpha_1 = \frac{1}{5} = 0,2 = \beta'_1, \quad \beta_1 = \frac{1}{5-0,2} = 0,20833 = \alpha'_1,$$

$$\alpha_2 = \frac{1}{5-0,20833} = 0,20869 = \beta'_2, \quad \beta_2 = \frac{1}{5-0,20869} = 0,20871 = \alpha'_2,$$

$$(\text{do } F_3) \alpha_2 = 0,20869, \quad \alpha_2^2 = 0,04355, \quad \alpha_2^3 = 0,00909, \quad \alpha_2^4 = 0,00190, \quad \alpha_2^5 = 0,00040,$$

$$(\text{do } F_6) \beta_2 = 0,20871, \quad \beta_2^2 = 0,04356, \quad \beta_2^3 = 0,00909, \quad \beta_2^4 = 0,00190, \quad \beta_2^5 = 0,00040,$$

$$(\text{do } F_8) \beta_1 = 0,20833, \quad \beta_1^2 = 0,04340, \quad \beta_1^3 = 0,00904, \quad \beta_1^4 = 0,00188, \quad \beta_1^5 = 0,00039,$$

$$(\text{do } F_1) \alpha_1 = 0,2, \quad \alpha_1^2 = 0,04, \quad \alpha_1^3 = 0,008, \quad \alpha_1^4 = 0,0016, \quad \alpha_1^5 = 0,00032$$

$$\frac{\beta_2 n}{r} = 0,00040, \quad \frac{rm}{D} = 0,00909, \quad \frac{\beta_2 n}{r} - \frac{rm}{D} = -0,00870,$$

$$\frac{\beta_1 p}{r} = 0,00868, \quad \frac{r}{D} = 0,2, \quad \frac{\beta_1 p}{r} - \frac{r}{D} = -0,19132,$$

$$\frac{\alpha_2 m}{D} = 0,00190, \quad \frac{\alpha_2 m}{D} - p = -0,03977,$$

$$T = \frac{r}{D} (m+1) = 0,20909, \quad \frac{\alpha_1}{D} - 1 = -0,96,$$

$$S = 1 + p + n = 1,04356, \quad S + T = 1,25266, \quad S - T = 0,83447.$$

Równania (42) dla warunków (a) przybierają postać

$$Q(-0,0087 F_6 - 0,19132 F_8 + 0,83447 - \cos \vartheta) + \\ + q'(-0,03977 F_3 - 0,96 F_1 - 0,0019 F_5 - 0,83447) = 0, \quad (42')$$

$$\psi_\vartheta = \frac{Q}{2} (-0,0087 F_6 - 0,19132 F_8 + 1,25266 - \cos \vartheta) + \\ + \frac{q'}{2} (-0,03977 F_3 - 0,96 F_1 - 0,0019 F_5 + 1,25266). \quad (43')$$

2. Sprawdzenie I'

$$\vartheta = 0, \quad P_1 = P_2 = P_3 = \dots = 1,$$

$$F_1 = 1 + 0,2 + 0,04 + 0,008 + \dots = 1,25,$$

$$F_6 = 1 + 0,20871 + 0,04356 + 0,00909 + 0,00190 + 0,00040 = 1,2637,$$

$$F_8 = 1 + 0,20833 + 0,04340 + 0,00904 + 0,00190 + 0,00040 = 1,2631,$$

$$F_3 = F_5 = F_6, \quad (46')$$

$$1 - (-0,0087 F_6 - 0,19132 F_8) = -(-0,03977 F_3 - 0,96 F_1 - 0,0019 F_5) = 1,25266,$$

$$1 + 0,0087 \cdot 1,2637 + 0,19132 \cdot 1,2631 = 0,03977 \cdot 1,2637 + 0,96 \cdot 1,25 +$$

$$+ 0,0019 \cdot 1,2637 = 1,25266,$$

$$1 + 0,0110 + 0,2417 = 0,0527 + 1,2 = 1,2527.$$

Sprawdzenie II

$$\vartheta = 180^\circ, \quad P_1 = P_3 = P_5 = \dots = -1, \quad P_2 = P_4 = P_6 = \dots = 1,$$

$$F_1 = 1 - 0,2 + 0,04 - 0,008 + 0,0016 - 0,0003 = 0,8333,$$

$$F_6 = 1 - 0,20871 + 0,04356 - 0,00909 + 0,00190 - 0,0004 = 0,8273,$$

$$F_8 = 1 - 0,20833 + 0,04340 - 0,00909 + 0,00190 - 0,0004 = 0,8275,$$

$$F_3 = F_5 = F_6, \quad (49')$$

$$1 + (-0,0087 F_6 - 0,19132 F_8) = -(-0,03977 F_3 - 0,96 F_1 - 0,0019 F_5) = 0,83477,$$

$$1 - 0,0087 \cdot 0,8273 - 0,19132 \cdot 0,8275 = 0,03977 \cdot 0,8273 + 0,96 \cdot 0,8333 -$$

$$- 0,0019 \cdot 0,8273 = 0,83477,$$

$$1 - 0,0072 - 0,1583 = 0,0345 + 0,8 = 0,8345.$$

3. Obliczenie kąta ϑ

$$\text{Zakładamy na przykład } Q = 2q'. \quad (b)$$

Pierwsza seria obliczeń ma charakter orientacyjny. Nie wiedząc zupełnie jaka wartość kąta ϑ odpowiada postawionym w (a) i (b) warunkom, przyjmujemy na razie dowolną wartość, np. $\vartheta = 90^\circ$. Wartości wielomianów P znajdujemy z tablic, np. z tablic Jahncke i Emde³. Ze względu na przybliżony charakter obliczeń orientacyjnych zamiast wzoru (43') przy tych obliczeniach możemy stosować skrócone równanie

$$Q(-0,009 F_6 - 0,191 F_8 + 0,834 - \cos \vartheta) + q'(-0,04 F_3 - 0,96 F_1 - 0,002 F_5 - 0,834) = 0$$

lub

$$-0,018 F_6 - 0,382 F_8 + 0,834 - 2 \cos \vartheta - 0,04 F_3 - 0,96 F_1 - 0,02 F_5 = 0.$$

³ Jahncke und Emde. *Funktionentafeln*. II wydanie niemieckie 1948. Wydanie rosyjskie 1949.

Wyniki naszych obliczeń są zatem

$$\begin{aligned} \text{przy } \vartheta = 103^\circ 30', \quad \Delta = -0,0141, \\ \vartheta = 105^\circ, \quad \Delta = +0,0427. \end{aligned}$$

Za pomocą interpolacji otrzymujemy

$$\vartheta = 103^\circ 30' + \frac{141}{568} \cdot 90 = 103^\circ 52'.$$

B. Sprawdzenie tożsamości (28) i (29).

Sprawdzenie algebraiczne tych tożsamości jest bardzo zawikłane. Ograniczymy się tu do sprawdzenia arytmetycznego przy założeniach:

$$1) r=r'=1, \quad D=5, \quad \vartheta=0 \quad \text{ i } \quad 2) r=r'=1, \quad D=5, \quad \vartheta=180.$$

1. Zakładając $\vartheta=0$ w (28) znajdujemy

$$\frac{rm}{D(r-\beta'_2)} + \frac{r'}{D(r-\beta'_1)} = \frac{1}{D-r} + \frac{p'}{D-r-a_1} + \frac{n'}{D-r-a_2}.$$

Podstawiając tu jeszcze $r=r'$ otrzymujemy tożsamość

$$\frac{rm}{D(r-a_2)} + \frac{r}{D(r-a_1)} = \frac{1}{D-r} + \frac{p}{D-r-\beta_1} + \frac{n}{D-r-\beta_2}.$$

Przyjmując $r=1$, $D=5$ i wykorzystując wyniki punktu A, mamy

$$\begin{aligned} \frac{0,04546}{5 \cdot 0,7913} + \frac{1}{5 \cdot 0,8} &= \frac{1}{4} + \frac{1}{24 \cdot 3,7917} + \frac{0,0019}{3,7913}, \\ 0,0115 + 0,25 &= 0,25 + 0,0110 + 0,0005. \end{aligned}$$

2. Mamy

$$\begin{aligned} \frac{n}{r-\beta_2} + \frac{p}{r-\beta_1} &= \frac{r}{D} \left(\frac{1}{D-r-a_1} + \frac{m}{D-r-a_2} \right), \quad r=1, \quad D=5, \\ \frac{0,0019}{0,7913} + \frac{0,0417}{0,7917} &= \frac{1}{5 \cdot 3,7917} + \frac{0,04546}{5 \cdot 3,7913}, \\ 0,0024 + 0,0528 &= 0,0528 + 0,0024. \end{aligned}$$

7. WARTOŚCI KĄTÓW GRANICZNYCH PRZY DANYCH POTENCJAŁACH

Obliczenia kątów granicznych przeprowadzono dla jednakowych kul ($r=r'$). Dla stosunku

$$\delta = \frac{D}{r} \quad (51)$$

przyjęto wartości $\delta = 3, 5, 10, 15, 20, 50$. Dla stosunku potencjałów obu kul

$$\kappa = \frac{V}{|V'|} = \frac{Q}{|q'|} \quad (52)$$

przyjęte zostały wartości $\kappa = 1,5; 2; 3; 4; 5$.

Przy bardzo wielkiej odległości obu kul, dla której usprawiedliwione jest założenie $\frac{D}{r} \rightarrow \infty$ otrzymujemy przybliżone wartości

$$\beta_2 = n = m = \beta_1 = p = \frac{r}{D} = \beta'_2 = p' = \beta'_1 = n' = 0, \quad F_1 = 1.$$

Po podstawieniu tych wartości do (42) znajdujemy równanie

$$Q(1 - \cos \vartheta) - 2q' = 0,$$

z którego obliczamy

$$\cos \vartheta = \frac{\kappa - 2}{\kappa}. \quad (53)$$

Wyrażenie powyższe możemy otrzymać bezpośrednio przeprowadzając następujące rozumowanie.

Przy bardzo dużej odległości D między kulami możemy przypuścić, że rozkład gęstości powierzchniowej σ ładunku elektrycznego na powierzchni kuli K jest równomierny. Wtedy całkowity strumień wychodzący z kuli K wynosi $Q = 4\pi r^2 \sigma$, strumień zaś przenikający czaszą kulistą odpowiadającą kątowi granicznemu ϑ i mającą wysokość h (por. rys. 1) jest $q' = 2\pi r h \sigma$. Wynika stąd

$$\frac{h}{2r} = \frac{1}{\kappa},$$

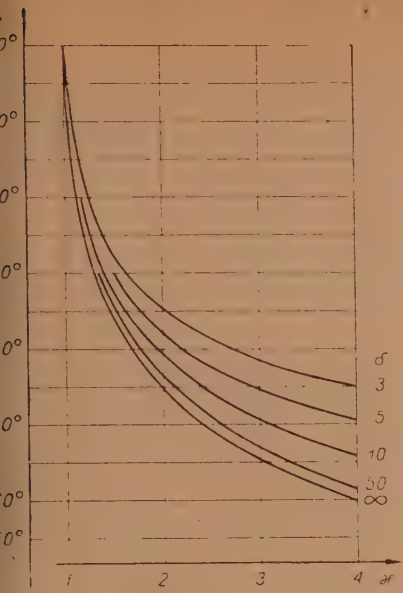
a ponieważ $h = r(1 - \cos \vartheta)$, przeto otrzymuje się tu znowu równanie (53)

Dla $\kappa = \infty$ z (42) otrzymuje się równanie

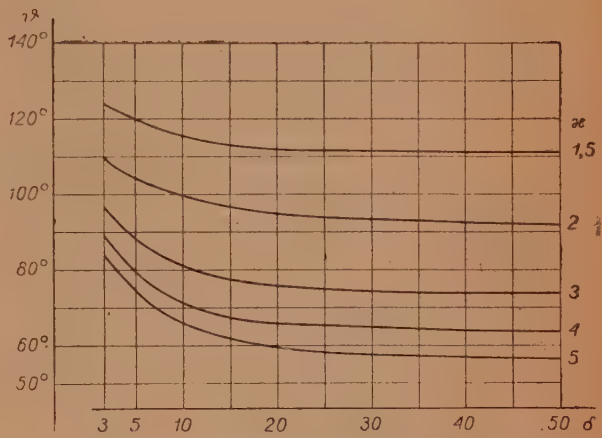
$$\left(\frac{\beta_2 n}{r} - \frac{r' m}{D}\right) F_6 + \left(\frac{\beta_1 p}{r} - \frac{r'}{D}\right) F_8 + 1 + p + n - \frac{r'}{D}(m+1) - \cos \vartheta = 0, \quad (54)$$

za pomocą którego możemy w zwykły sposób obliczać ϑ . Dla $\kappa = 1$ jest $\vartheta = 180^\circ$ (por. rozdz. 5, sprawdzenie II). Dla $\delta = \infty$ otrzymujemy $\vartheta = 0$; ta sama wartość kąta ϑ wynika z (53) po podstawieniu $\kappa = \infty$.

Wyniki obliczeń przedstawiono w tablicy 1 oraz na rys. 10 i 11. Szczególnie interesujące są dane dla $\kappa = \infty$ wskazujące na to, że w żadnej konfiguracji geometrycznej nie możemy przez zwiększenie κ dowolnie zmniejszyć wartości kąta granicznego. (Przypadek $\kappa = \infty$ można osiągnąć przez uziemienie kuli K').



Rys. 10. Wartości kątów granicznych w układzie dwóch kul przy danych potencjałach.



Rys. 11. Wartości kątów granicznych w układzie dwóch kul przy danych potencjałach.

Tablica 1

Wartości kątów granicznych δ w układzie 2 kul przy danych potencjałach

$\delta \backslash \sigma$	1	1,5	2	3	4	5	∞
3	180°	123°44'	109°44'	96°21'	89°17'	85°1'	66°50'
5	180°	120°	103°52'	88°	80°16'	73°45'	51°2'
10	180°	115°47'	98°6'	80°21'	71°14'	65°35'	36°49,5'
15	180°	113°37'	95°21'	77°18'	67°44'	61°38'	29°53'
20	180°	112°39'	94°5'	75°41'	65°54'	59°37'	25°54'
50	180°	110°48'	91°41'	72°39'	62°28'	55°50'	16°19'
∞	180°	109°28'	90°	70°32'	60°	53°80'	0

8. KĄTY GRANICZNE NA KULI WYWOŁANE PRZEZ ŁADUNEK PUNKTOWY

W przypadku, gdy promień kuli K' jest znacznie mniejszy od promienia kuli K , w przybliżeniu możemy założyć

$$r' = 0 . \tag{55}$$

Konsekwencją tego założenia jest

$$\alpha_i = 0, \quad \beta_i = \beta = \frac{r^2}{D}, \quad \frac{1}{k} = 0, \quad p = p' = 0, \quad (56)$$

$$R = 1, \quad n = n' = 0, \quad m = 0, \quad S = 1, \quad S' = 1, \quad T = 0, \quad T' = \frac{r}{D}.$$

Podstawiając powyższe wartości do (42) otrzymujemy równanie

$$Q(1 - \cos \vartheta) + q' \left[\left(\frac{r^2}{D^2} - 1 \right) F_1 - 1 + \frac{r}{D} \right] = 0, \quad (57)$$

za pomocą którego można obliczać kąty graniczne w rozważanym przypadku.

Zakładając w (57) $r = 0$ znajdujemy, że względu na $F_1 = 1$, zależność

$$1 - \cos \vartheta = 2 \frac{q'}{Q}$$

prowadzącą do znanego już związku (53). Przyczyna tej zgodności jest oczywista.

Skoro mamy obecnie do czynienia z ładunkiem punktowym, trudno jest porównywać ze sobą potencjały V i V' (potencjał V' jest nieokreślony) lub proporcjonalne do nich wielkości q' i Q . Jako wielkości odniesienia musimy w rozważanym przypadku przyjąć całkowite ładunki na K i K' . Wynoszą one (patrz [A 95 i 96])

$$H = Q + \frac{r}{D} q', \quad H' = q'. \quad (58)$$

Oznaczając stosunek tych wielkości przez χ

$$\chi = \frac{H}{H'} \quad (59)$$

znajdujemy z (57) równanie służące do obliczania ϑ

$$\left(\chi - \frac{r}{D} \right) (1 - \cos \vartheta) - \left(1 - \frac{r}{D} \right) \left[\left(\frac{r}{D} + 1 \right) F_1 + 1 \right] = 0. \quad (60)$$

Dla $\chi = 1$ (oba ładunki jednakowe) równanie (60) powinno być spełnione przez $\vartheta = 180^\circ$. Istotnie, mamy w tym przypadku

$$P_1 = -1, \quad P_2 = +1, \quad P_3 = -1, \quad P_4 = +1 \dots,$$

zatem

$$F_1 = 1 - \frac{r}{D} + \frac{r^2}{D^2} - \frac{r^3}{D^3} + \dots$$

Z drugiej strony z (60) wynika

$$2 = \left(\frac{r}{D} + 1 \right) F_1 + 1$$

co daje zgodny wynik.

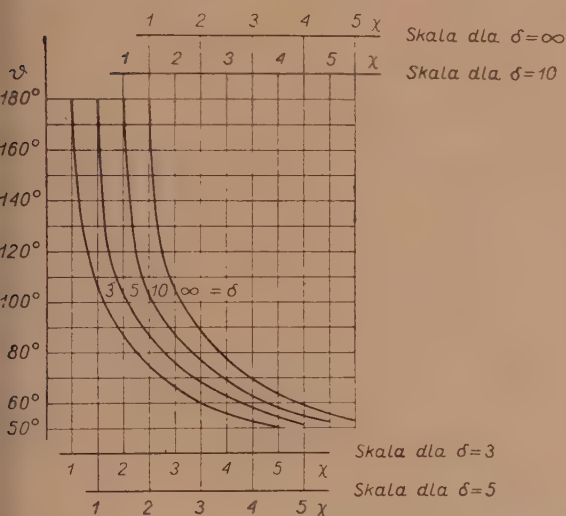
W przypadku gdy $\frac{D}{r} \rightarrow \infty$, to $P_1 \rightarrow 1$; z równania (60) wynika wówczas

$$\cos \vartheta = \frac{\chi - 2}{\chi}, \quad (61)$$

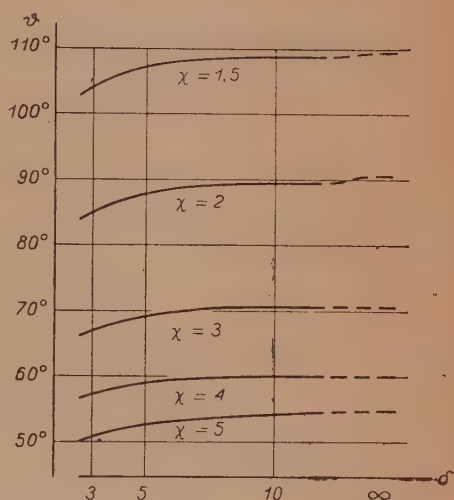
czyli związek analogiczny do (53).

Zakładając w (60) $\chi \rightarrow \infty$ otrzymujemy $\vartheta \rightarrow 0$.

Wartości kątów ϑ w rozważanym przypadku podane zostały w tabelicy 2 i przedstawione na rys. 12 i 13.



Rys. 12. Wartości kątów granicznych w układzie ładunek punktowy — kula (przy danych ładunkach).



Rys. 13. Wartości kątów granicznych w układzie ładunek punktowy — kula (przy danych ładunkach).

Tabela 2

Wartości kątów granicznych w układzie: ładunek punktowy — kula

$\sigma \backslash \chi$	1	1,5	2	3	4	5	∞
3	180°	104°15'	85°18'	67°6'	57°24'	50°22'	0
5	180°	107°31'	88°18'	69°21'	59°8'	52°30'	0
10	180°	108°58'	89°32'	70°15'	59°48'	52°59'	0
∞	180°	109°28'	90°	70°32'	60°	53°8'	0

Widzimy, że przy zwiększaniu ładunku kuli K (i przy stałym δ) zmniejsza się kąt graniczny ϑ ; ϑ zmniejsza się również przy zmniejszaniu odległości δ i zachowaniu stałych ładunków na K i K' .

9. KĄTY GRANICZNE PRZY DANYCH ŁADUNKACH

W celu uproszczenia obliczeń wprowadzamy następujące oznaczenia:

$$\left(\frac{\beta_2 n}{r} - \frac{r' m}{D}\right) F_6 + \left(\frac{\beta_1 p}{r} - \frac{r'}{D}\right) F_8 - \cos \vartheta = A_\vartheta, \quad (62)$$

$$\left(\frac{\beta_2 m}{D} - p'\right) F_3 + \left(\frac{\beta_1'}{D} - 1\right) F_1 - n' F_5 = B_\vartheta, \quad (63)$$

$$SS' - TT' = U. \quad (64)$$

Równania (42) i (43) przybierają wtedy postać

$$\Delta = Q(A_\vartheta + S - T) + q'(B_\vartheta - S' + T') = 0, \quad (65)$$

$$\psi_\vartheta = \frac{Q}{2} (A_\vartheta + S + T) + \frac{q'}{2} (B_\vartheta + S' + T'), \quad (66)$$

Otrzymane na podstawie (21) i (22) wartości

$$Q = \frac{S'H - T'H'}{SS' - TT'}, \quad q' = \frac{SH' - TH}{SS' - TT'} \quad (67)$$

podstawiamy do (65) oraz (66) i znajdujemy równanie do obliczania kąta ϑ

$$\Delta = H(S'A_\vartheta - TB_\vartheta + U) - H'(T'A_\vartheta - SB_\vartheta + U) = 0 \quad (68)$$

oraz równanie do obliczania strumienia

$$\psi_\vartheta = \frac{H}{2U} (S'A_\vartheta - TB_\vartheta) + \frac{H'}{2U} (SB_\vartheta - T'A_\vartheta) + \frac{1}{2} (H + H'). \quad (69)$$

Metody sprawdzania powyższych równań są analogiczne do metod przedstawionych w rozdz. 5.

Sprawdzenie I.

Dla $\vartheta = 0$ musi być $\psi_{\vartheta=0} = 0$. Wynikają stąd warunki

$$TB_{\vartheta=0} - S'A_{\vartheta=0} = T'A_{\vartheta=0} - SB_{\vartheta=0} = U \quad (70)$$

albo przy $r = r'$

$$A_{\vartheta=0} = B_{\vartheta=0} = S + T. \quad (71)$$

Sprawdzenie II.

Dla $\vartheta = \pi$ musi być $\psi_{\vartheta=\pi} = H = H'$. Wynika stąd warunek

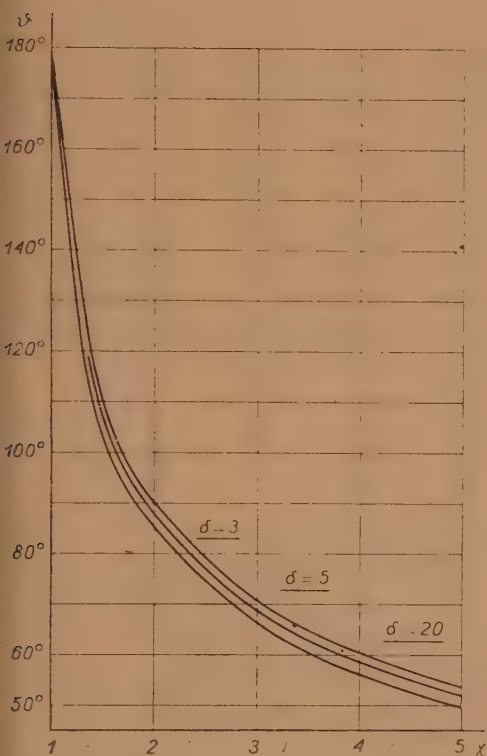
$$A_\vartheta(S' - T') + B_\vartheta(S - T) = 0 \quad (72)$$

albo przy $r=r'$

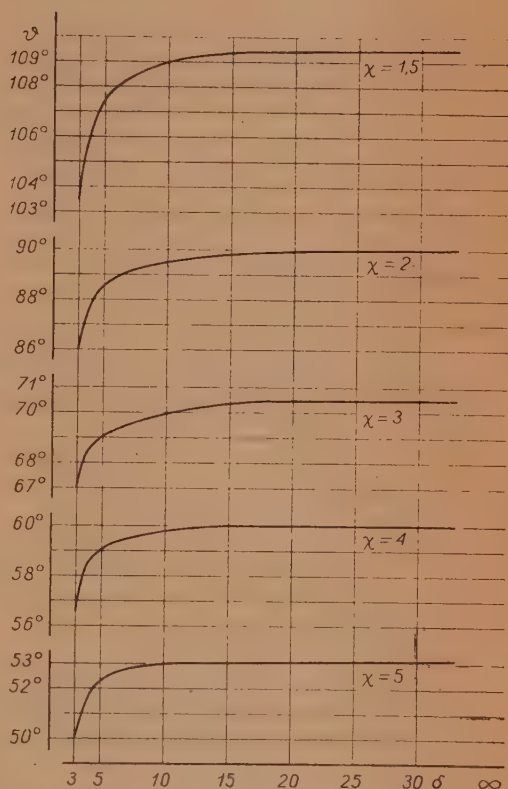
$$A_{\vartheta=\pi} = -B_{\vartheta=\pi} \approx S - T \quad (\text{por. rozdz. 5}). \quad (73)$$

Na podstawie powyższych wzorów obliczone zostały kąty graniczne dla rozmaitych wartości parametrów δ i χ . Obliczenia wykonano za pomocą metody przedstawionej już w rozdz. 6, jednakże tutaj rachunki są znacznie dłuższe.

Wyniki obliczeń przedstawione są w tablicy 3 oraz na rys. 14 i 15.



Rys. 14. Wartości kątów granicznych w układzie dwóch kul o danych ładunkach niejednakowych znaków.



Rys. 15. Wartości kątów granicznych w układzie dwóch kul o danych ładunkach niejednakowych znaków.

Szczególnie interesujący jest tu fakt, że przy bardzo dużych odległościach obu kul, kąty ϑ dążą do określonej granicy (patrz wzór 61). Ciekawe jest również porównanie rys. 15 z rys. 11. Z porównania wynika, że przebieg zmiany kątów granicznych przy stałych potencjałach wykazuje tendencję opadającą, podczas gdy przebieg zmiany kątów granicznych przy stałych ładunkach ma tendencję wzrastającą (to samo zjawisko obserwowaliśmy w rozdz. 8 przy ładunkach punktowych).

Tablica 3

Wartości kątów granicznych ϑ w układzie dwóch kul przy danych ładunkach

$\delta \backslash \chi$	1	1,5	2	3	4	5	∞
3	180°	103°32'	85°40'	66°43'	56°21'	49°59'	0
5	180°	107°28'	88°35'	69°19'	59°4'	52°24'	0
10	180°	108°56'	89°26'	69°44'	59°47'	52°58'	0
15	180°	109°15'	89°49'	70°23'	59°56'	53°4'	0
20	180°	109°21'	89°54'	70°26'	59°58'	53°6'	0
50	180°	109°27'	89°59'	70°31'	60°	53°8'	0
∞	180°	109°28'	90°	70°32'	60°	53°8'	0

Z fizycznego znaczenia kąta granicznego wynika, że dla $\chi = \infty$ musi być $\vartheta = 0$. Dla tego przypadku na podstawie wzoru (68) otrzymuje się

$$S'A_{\vartheta} - TB_{\vartheta} + U = 0.$$

Jak widać z tożsamości (70) równość powyższa jest spełniona przy $\vartheta = 0$.

10. OCENA BŁĘDÓW STOSOWANEJ METODY

Przedstawiona w pracy metoda jest metodą przybliżoną. Przybliżenie uwarunkowane jest tym, że:

1. Rzeczywistą sumę ładunków w przestrzeniach kulistych K' i K przedstawiamy za pomocą przybliżonych wyrażeń S , T , S' , T' .

2. Nieskończony ciąg ładunków wewnątrz przestrzeni K' lub K zastępujemy przez 5 ładunków (patrz wyżej rozdz. 4 i rys. 7), przy czym 4 ładunki mają wartości wyznaczone teoretycznie, piąty zaś ładunek ma charakter ładunku zastępczego; wartość jego równa się różnicy między łącznym ładunkiem całej kuli i sumą pierwszych 4 ładunków użytych do obliczenia. Poza tym pierwsze 4 ładunki znajdują się — przy przedstawionym sposobie obliczenia — w miejscach wyznaczonych przez teorię. Przyjmuje się jednakże, że suma pozostałych ładunków (tj. wspomniany wyżej ładunek zastępczy) występuje łącznie w miejscu, które teoria określa tylko dla piątego ładunku.

Dwa powyższe przybliżenia są źródłem błędów. Trzecie źródło błędów można by upatrywać w równaniu (32), którego rozwiązanie z natury rzeczy może być tylko przybliżone.

Górną granicę błędu popełnianego przez przybliżenie 1. można łatwo ustalić; zrobimy to niżej. O wiele trudniejsze jest stwierdzenie, jaka jest dokładność przy zastosowaniu przybliżenia 2. Ocenę popełnianego tu błędu dokonamy przez przeprowadzenie ścisłego obliczenia w pewnym przypadku charakterystycznym, co pozwoli wyciągnąć wnioski ogólniejsze. Nie będziemy natomiast rozważali wpływu trzeciego źródła błędów, ponieważ teoretycznie jesteśmy w stanie rozwiązać równanie (32) z dowolnąadaną z góry dokładnością.

Obliczenie błędu popełnionego przy przybliżeniu 1.

Obliczenie to opieramy na wywodach z [A rozdz. 7—G i H]. Rzeczywista wartość stosunku wynosi według [A 72]

$$\frac{Q_{n+1}}{Q_n} = z_2(1 - B_{Qn}), \quad (74)$$

przy czym $n=2,3,\dots$ i B_{Qn} jest funkcją n . Wszystkie B_{Qn} są liczbami dodatnimi. W ciągu $B_{Q2}, B_{Q3}, \dots$ największą wartość $B_{Q \max}$ ma wyrażenie B_{Q2} . Załóżmy, że błąd rachunku jest większy niż popełniany rzeczywiście i przyjmijmy, że stosujemy wzór rekurencyjny

$$\frac{Q_{n+1}}{Q_n} = z_2(1 - B_{Q \max}) = \xi \quad n=1, 2, 3, \dots, \quad (75)$$

gdzie

$$z_2 = \frac{D^2 - r^2 - r'^2}{2rr'} - \sqrt{\left(\frac{D^2 - r^2 - r'^2}{2rr'}\right)^2 - 1}. \quad (76)$$

por. [A 51].

Ciąg ładunków Q, Q_i jest

$$Q, Q_1 = Qp, Q_2^*, Q_3 = Q_1\xi^2, Q_4 = Q_1\xi^3, \dots, Q_n = Q_1\xi^{n-1}, \dots$$

Zamiast ścisłej wartości (wynikającej z [A 38] i [A 52])

$$Q_2^* = Q \frac{p}{k-p}, \quad (77)$$

gdzie

$$p = \frac{rr'}{D^2 - r'^2}, \quad k = \frac{D^2 - r^2 - r'^2}{rr'}, \quad (78)$$

przy obliczaniu sumy S przyjmujemy wartość przybliżoną

$$Q_2 = Q_1 z_2, \quad (79)$$

popełniając przy tym błąd dodatkowy

$$Q_2^* - Q_2 = Q\beta_Q. \quad (80)$$

Dla β_Q z odnośnych wzorów w przypadku $r=r'$ otrzymujemy wyrażenie

$$\beta_Q = \frac{1}{\delta^2 - 1} \left[\frac{1}{\delta^2 - 2 - (\delta^2 - 1)^{-1}} - \frac{1}{2} (\delta^2 - 2) + \sqrt{\left(\frac{\delta^2 - 2}{2} \right)^2 - 1} \right]. \quad (81)$$

Obliczając wartość β_Q dla rozmaitych wartości δ , znajdujemy

δ	3	4	5
β_Q	0,00006	0,000002	0,0000 ...

Widać stąd, że przyjmując wartość Q_2 zamiast Q_2^* popełniamy błąd bardzo mały. Wielkość β_Q jest dodatnia, a zatem

$$Q_2 > Q_2^*.$$

Rozważmy teraz trzy szeregi: szereg ścisły QS^* odpowiadający rzeczywistemu ładunkowi kuli K

$$QS^* = Q + Q_1 + Q_2^* + \sum_{n=2}^{\infty} [Q_1 z_2^n \prod_{k=1}^n (1 - B_{Q_k})], \quad (82)$$

szereg QS' otrzymany z QS^* przez zamianę B_{Q_k} na $B_{Q_{\max}}$

$$\begin{aligned} QS' &= Q + Q_1 + Q_2^* + \sum_{n=2}^{\infty} Q_1 z_2^n (1 - B_{Q_{\max}})^n = \\ &= Q + Q_1 + Q_1 z_2 + Q \beta_Q + Q_1 \sum_{n=2}^{\infty} z_2^n (1 - B_{Q_{\max}})^n = \\ &= Q + Q \beta_Q + Q_1 z_2 B_{Q_{\max}} + \frac{Q_1}{1 - z_2 (1 - B_{Q_{\max}})} \end{aligned} \quad (83)$$

oraz szereg QS stosowany przy obliczeniach

$$QS = Q + Q_1 \sum_{n=1}^{\infty} z_2^{n-1} = Q + \frac{Q_1}{1 - z_2} = Q + \frac{Qp}{1 - z_2}. \quad (84)$$

Stwierdzamy przy tym, że

$$S' < S^* < S + \beta_Q. \quad (85)$$

Rzeczywisty błąd względny popełniany przy stosowaniu S zamiast S^* , oznaczamy przez

$$s = \frac{S - S^*}{S}. \quad (86)$$

Mamy teraz

$$s < \frac{S - S^* + \beta_Q}{S} < \frac{S - S' + \beta_Q}{S'} = s'. \quad (87)$$

Dla s' znajdujemy wyrażenie

$$s' = p \frac{\frac{1}{1-z_2} - z_2 B_{Q \max} - \frac{1}{1-z_2(1-B_{Q \max})}}{1 + \frac{p}{1-z_2}}, \quad (88)$$

gdzie (w przypadku $r=r'$)

$$p = \frac{1}{\delta^2 - 1}, \quad z_2 = \frac{k}{2} - \sqrt{\frac{k^2}{4} - 1}, \quad k = \delta^2 - 2, \quad B_{Q \max} = z_2^4. \quad (89)$$

Podstawiając powyższe wyrażenia znajdujemy, że górny kres s' błędu s jest liczbą nadzwyczaj małą, mianowicie:

δ	3	4	5
p	0,125	0,06667	0,04167
z_2	0,145899	0,071797	0,043561
S	1,14635	1,07183	1,043568
$S - S' + \beta_Q$	0,000003	0,000 ...	0,000 ...
s'	0,0000026	0,000 ...	0,000 ...

W podobny sposób można obliczyć górny kres błędu wyrażenia T . Rzeczywiste ładunki w przestrzeni K' tworzą ciąg

$$q_1, q_2, q_3^*, q_4^*, q_5^* \dots,$$

którego sumę wyrazów można przedstawić w postaci

$$T^* = q_1 + q_2 + q_3^* + \sum_{n=3}^{\infty} \left[q_2 z_2^{n-1} \prod_{k=4}^n (1 - B_{q_k}) \right]. \quad (90)$$

Z [A 38 i A 59] wynika, że

$$q_3^* = \frac{Q}{\delta(k^2 - 1)}, \quad q_2 = \frac{Q}{\delta k}, \quad q_1 = \frac{Q}{\delta}. \quad (91)$$

Przy obliczaniu sumy T wyraz q_3^* zastępujemy przez

$$q_3 = q_2 z_2 = \frac{Q z_2}{\delta k} \quad (92)$$

i w wyniku

$$q_3 - q_3^* = \frac{Q}{\delta} \left(\frac{z_2}{k} - \frac{1}{k^2 - 1} \right) = \beta_Q Q. \quad (93)$$

Po podstawieniu wartości liczbowych znajdujemy

δ	3	4	5
β_Q	$3 \cdot 10^{-6}$	$25 \cdot 10^{-8}$	$2 \cdot 10^{-8}$

Rozważmy także szereg

$$QT' = q_1 + q_2 + q_3^* + \sum_{n=3}^{\infty} q_2 z_2^{n-1} (1 - B_{q \max})^{n-1} \quad (94)$$

w którym zgodnie z [A str. 89] zakładamy

$$B_{q \max} = z_2^6. \quad (95)$$

Szereg ten możemy napisać jako

$$\begin{aligned} QT' &= q_1 + q_2 + q_3 - \beta_q Q + q_2 \sum_{n=3}^{\infty} z_2^{n-1} (1 - B_{q \max})^{n-1} = \\ &= q_1 - \beta_q Q + q_2 z_2 B_{q \max} + \frac{q_2}{1 - z_2 (1 - B_{q \max})}. \end{aligned}$$

Do obliczeń używamy trzeciego szeregu,

$$QT' = q_1 + q_2 + q_2 z_2 + q_2 z_2^2 + \dots = q_1 + \frac{q_2}{1 - z_2}. \quad (96)$$

Mamy tu

$$T' < T^* < T - \beta, \quad (97)$$

błąd zaś względny t wynosi

$$t = \frac{T - T^*}{T} < \frac{T - T'}{T} = t'. \quad (98)$$

W celu obliczenia górnego kresu t' błędu znajdujemy

$$T - T' = \frac{1}{\delta k} \left[\frac{1}{1 - z_2} - z_2 B_{q \max} - \frac{1}{1 - z_2 (1 - B_{q \max})} \right] + \beta_q.$$

Ze względu na nadzwyczaj małą wartość $B_{q \max}$ czynnik znajdujący się w nawiasie kwadratowym bardzo mało różni się od zera (ma on wartość dodatnią), wobec tego możemy w przybliżeniu założyć

$$T - T' \approx \beta_q. \quad (99)$$

Korzystając z tablicy [A str. 102] znajdujemy

δ	3	4	5
$t' = \frac{\beta_q}{T}$	0,000008	0,000001	0,0000001

Jak widać z powyższych obliczeń wartości s' i t' są nadzwyczaj małe; wynika stąd, że wartości współczynników obliczane za pomocą wzorów (84) i (96), są bardzo dokładne⁴ (można powiedzieć: zadziwiająco dokładne).

⁴ Przy obliczeniach praktycznych dokładność powyższych wzorów jest zazwyczaj znacznie mniejsza ze względu na konieczność ograniczenia rachunku do niewielkiej liczby znaków dziesiętnych oraz wskutek korzystania z tablic funkcji hiperbolicznych o ograniczonej dokładności; jednak i tu stosunkowo łatwo dla s' i t' można osiągnąć wartości 0,0001.

Obliczenie błędu popełnionego przy przybliżeniu 2.

W celu oceny błędu wprowadzonego przez opisane wyżej przybliżenie drugie przeprowadzimy raz jeszcze obliczenie kąta granicznego w najniekorzystniejszych warunkach przy założeniu określonego stopnia dokładności obliczenia. Najmniej korzystnym układem (tzn. prowadzącym do stosunkowo największych błędów) jest układ, w którym $\frac{D}{r} = \delta = 3$ (zakładamy, że $r = r'$). Obliczenie przeprowadzimy tak, aby w granicach pięciu znaków po przecinku otrzymać dokładne wyniki.

W celu spełnienia ostatniego warunku będziemy musieli (jak wykazemy niżej) uwzględnić nie jak dotychczas 10 ładunków punktowych (5 wewnątrz przestrzeni K' i 5 — wewnątrz K), lecz 18 ładunków punktowych (9 wewnątrz przestrzeni K' i 9 — wewnątrz K). Przyjmując te same oznaczenia co poprzednio możemy wykonać następujące zestawienie:

Punkt wewnątrz K' , w którym umieszczony jest ładunek	Odległość tego punktu od środka M' kuli K'	Wielkość ładunku punktowego	Punkt wewnątrz K , w którym umieszczony jest ładunek	Odległość tego punktu od środka M kuli K	Wielkość ładunku punktowego
M'	0	$-q'$	M	0	Q
A_1	a_1	$-q_1$	B_1	β_1	Q_1
A_1'	a_1'	$-q_1'$	B_1	β_1	Q_1
A_2	a_2	$-q_2$	B_2'	β_2'	Q_2'
A_2'	a_2'	$-q_2'$	B_2	β_2	Q_2
A_3	a_3	$-q_3$	B_3'	β_3'	Q_3'
A_3'	a_3'	$-q_3'$	B_3	β_3	Q_3
A_4	a_4	$-q_4$	B_4'	β_4'	Q_4'
A_4'	a_4'	$-q_4'$	B_4	β_4	Q_4

Podobnie jak we wzorach (23) i (24) mamy tu

$$\alpha_n = \beta_n' \quad \text{i} \quad \alpha_n' = \beta_n. \quad (100)$$

Wartości $\alpha_n, \beta_n, Q_n, Q_n', q_n$ i q_n' obliczymy nie posilując się przybliżeniem, lecz ścisłymi wzorami [A 38]; w celu łatwego przedstawienia rezultatów tych obliczeń wprowadzimy następujące oznaczenia:

$$\begin{aligned} \delta^2 - 1 &= C_2, \\ \delta^3 - 2\delta &= C_3, \\ \delta^4 - 3\delta^2 + 1 &= C_4, \end{aligned} \quad (101)$$

$$\begin{aligned}
 \delta^5 - 4\delta^3 + 3\delta &= C_5, \\
 \delta^6 - 5\delta^4 + 6\delta^2 - 1 &= C_6, \\
 \delta^7 - 6\delta^5 + 10\delta^3 - 4\delta &= C_7, \\
 \delta^8 - 7\delta^6 + 15\delta^4 - 10\delta^2 + 1 &= C_8.
 \end{aligned}
 \tag{101}$$

Możemy teraz napisać

$$\begin{aligned}
 a_1 &= \frac{1}{\delta}, & q_1 &= \frac{Q}{\delta}, & Q_1 &= \frac{q'}{\delta}, \\
 a'_1 &= \frac{\delta}{C_2}, & q'_1 &= \frac{q'}{C_2}, & Q_1 &= \frac{Q}{C_2}, \\
 a_2 &= \frac{C_2}{C_3}, & q_2 &= \frac{Q}{C_3}, & Q'_2 &= \frac{q'}{C_3}, \\
 a'_2 &= \frac{C_3}{C_4}, & q'_2 &= \frac{q'}{C_4}, & Q_2 &= \frac{Q}{C_4}, \\
 a_3 &= \frac{C_4}{C_5}, & q_3 &= \frac{Q}{C_5}, & Q'_3 &= \frac{q'}{C_5}, \\
 a'_3 &= \frac{C_5}{C_6}, & q'_3 &= \frac{q'}{C_6}, & Q_3 &= \frac{Q}{C_6}, \\
 a_4 &= \frac{C_6}{C_7}, & q_4 &= \frac{Q}{C_7}, & Q'_4 &= \frac{q'}{C_7}, \\
 a'_4 &= \frac{C_7}{C_8}, & q'_4 &= \frac{q'}{C_8}, & Q_4 &= \frac{Q}{C_8}.
 \end{aligned}
 \tag{102}$$

Podstawiając do powyższych wyrażeń $\delta=3$, znajdujemy wartości

1. $a_1 = \beta'_1 = 0,33333$	$q_1 = 0,33333 \ Q$	$Q'_1 = 0,33333 \ q'$
2. $a'_1 = \beta_1 = 0,37500$	$q'_1 = 0,12500 \ q'$	$Q_1 = 0,12500 \ Q$
3. $a_2 = \beta'_2 = 0,38095$	$q_2 = 0,04762 \ Q$	$Q'_2 = 0,04762 \ q'$
4. $a'_2 = \beta_2 = 0,38182$	$q'_2 = 0,01818 \ q'$	$Q_2 = 0,01818 \ Q$
5. $a_3 = \beta'_3 = 0,38194$	$q_3 = 0,00694 \ Q$	$Q'_3 = 0,00694 \ q'$
6. $a'_3 = \beta_3 = 0,38196$	$q'_3 = 0,00265 \ q'$	$Q_3 = 0,00265 \ Q$
7. $a_4 = \beta'_4 = 0,38197$	$q_4 = 0,00101 \ Q$	$Q'_4 = 0,00101 \ q'$
8. $a'_4 = \beta_4 = 0,38197$	$q'_4 = 0,00039 \ q'$	$Q_4 = 0,00039 \ Q$

Obliczając a_n i a'_n dla $n > 4$ stwierdzamy, że odnośne wartości minimalnie tylko przekraczają wartość a_4 . Możemy zatem, rozważając tylko 5 znaków dziesiętnych, przyjąć

$$a_n = a'_n = 0,38197.$$

Wynika stąd, że wszystkie ładunki q_n, q'_n, Q_n, Q'_n z indeksem n większym od 4 możemy umiejscowić w punktach A_4 lub B_4 . Ładunki te są wyrazami malejącego postępu geometrycznego o mnożniku $0,1459 = 0,38197^2$. Suma ładunków pochodzących od Q i położonych wewnątrz K wynosi zatem

$$Q \left(1 + 0,125 + 0,01818 + 0,00265 + 0,00039 \frac{1}{1 - 0,1459} \right) = \\ = Q(1 + 0,125 + 0,01818 + 0,00265 + 0,00046) = 1,14629 Q.$$

W podobny sposób znajdujemy sumę ładunków pochodzących od Q i położonych wewnątrz K'

$$Q \left(0,33333 + 0,04762 + 0,00694 + 0,00101 \frac{1}{1 - 0,1459} \right) = \\ = Q(0,33333 + 0,04762 + 0,00694 + 0,00118) = 0,38907 Q.$$

Wynika stąd, że w celu uwzględnienia ładunków z indeksami większymi od 4 musimy skorygować dwa dolne wiersze ostatniej tablicy. Przybierają one postać:

7'	$\alpha_4 = \beta'_4 = 0,38197$	$q_4 = 0,00118 Q$	$Q'_4 = 0,00118 q'$
8'	$\alpha'_4 = \beta_4 = 0,38197$	$q'_4 = 0,00046 q'$	$Q_4 = 0,00046 Q$

Dla porównania podajemy podobną tablicę, która w rozdz. 6÷8 służyła do obliczania kątów granicznych przy $\delta = 3$.

$\alpha_1 = \beta'_1 = 0,33333$	$q_1 = 0,33333 Q$	$Q'_1 = 0,33333 q'$
$\alpha'_1 = \beta_1 = 0,37500$	$q'_1 = 0,12500 q'$	$Q_1 = 0,12500 Q$
$\alpha_2 = \beta'_2 = 0,38095$	$q_2 = 0,05576 Q$	$Q'_2 = 0,05576 q'$
$\alpha'_2 = \beta_2 = 0,38192$	$q'_2 = 0,02135 q'$	$Q_2 = 0,02135 Q$

Sumy ładunków wynosiły tu $0,38909 Q$ i $1,14635 Q^5$

Musimy zestawić obecnie równania analogiczne do (26÷38), ale dostosowane do 18 ładunków zastępczych. W celu uproszczenia ich wprowadzamy następujące skróty:

⁵ Na zasadzie rozważań w punkcie 1 należy sądzić, że liczba ta jest bliższa rzeczywistości niż liczba 1,14629 otrzymana z rachunku *dokładniejszego*.

$$\frac{Q_i}{Q} = \frac{q'_i}{q'} = K_i, \quad \frac{q_i}{Q} = \frac{Q'_i}{q'} = k_i, \quad \frac{Q}{q'} = \kappa. \quad (103)$$

Równanie odpowiadające równaniu (26) jest obecnie

$$\psi_\vartheta = \frac{q'}{2} \left\{ - \left(\frac{D - r \cos \vartheta}{\varrho'_M} - 1 \right) + \kappa (1 - \cos \vartheta) + \sum_{i=1}^4 \left[-k_i \kappa \left(\frac{D - a_i - r \cos \vartheta}{\varrho_{ai}} - 1 \right) - K_i \left(\frac{D - a'_i - r \cos \vartheta}{\varrho_{ai}} - 1 \right) + K_i \kappa \left(\frac{\beta_i - r \cos \vartheta}{\varrho_{\beta i}} + 1 \right) + k_i \left(\frac{\beta'_i - r \cos \vartheta}{\varrho_{\beta i}} + 1 \right) \right] \right\}. \quad (104)$$

Zamiast zależności (27) znajdujemy

$$\frac{1}{4\pi\epsilon_0} \left[-\frac{1}{\varrho'_M} + \frac{\kappa}{r} + \sum_{i=1}^4 \left(-\frac{k_i \kappa}{\varrho_{ai}} - \frac{K_i}{\varrho'_{ai}} + \frac{K_i \kappa}{\varrho_{\beta i}} + \frac{k_i}{\varrho'_{\beta i}} \right) \right] = \frac{\kappa}{4\pi\epsilon r}. \quad (105)$$

Współczynnik κ może być obrany dowolnie. Wynikają stąd zatem dwie zależności odpowiadające równościom (28) i (29)

$$-\frac{1}{\varrho'_M} + \sum_{i=1}^4 \left(-\frac{K_i}{\varrho'_{ai}} + \frac{k_i}{\varrho_{\beta i}} \right) = 0, \quad \sum_{i=1}^4 \left(-\frac{k_i}{\varrho_{ai}} + \frac{K_i}{\varrho_{\beta i}} \right) = 0. \quad (106)$$

Uwzględniając te związki znajdujemy równanie odpowiadające równaniom (30) i (31)

$$\psi_\vartheta = \frac{q'}{2} \left[-\frac{D}{\varrho'_M} + 1 + \kappa - \kappa \cos \vartheta + \sum_{i=1}^4 \left(-k_i \kappa \frac{D - a_i}{\varrho_{ai}} - K_i \frac{D - a'_i}{\varrho'_{ai}} + K_i \kappa \frac{\beta_i}{\varrho_{\beta i}} + k_i \frac{\beta'_i}{\varrho'_{\beta i}} + k_i \kappa + K_i + K_i \kappa + k_i \right) \right] = q' \left[1 + \sum_{i=1}^4 (k_i \kappa + K_i) \right]. \quad (107)$$

Następnie otrzymujemy równanie służące do obliczenia kąta ϑ (odpowiednik równania 32)

$$\Delta = \frac{q'}{2} \left[-\frac{D}{\varrho'_M} - 1 + \kappa - \kappa \cos \vartheta + \sum_{i=1}^4 \left(-k_i \kappa \frac{D - a_i}{\varrho_{ai}} - K_i \frac{D - a'_i}{\varrho'_{ai}} + K_i \kappa \frac{\beta_i}{\varrho_{\beta i}} + k_i \frac{\beta'_i}{\varrho'_{\beta i}} - k_i \kappa - K_i + K_i \kappa + k_i \right) \right] = 0. \quad (108)$$

Sprawdzimy wartości kątów ϑ , obliczonych poprzednio metodą przybliżoną (patrz tablica 1) w trzech przypadkach

- $q' = 0, \quad \vartheta = 66^\circ 50',$
- $Q = 1,5 q', \quad \vartheta = 123^\circ 44',$
- $Q = q', \quad \vartheta = 180^\circ.$

Obliczenia wykonamy bezpośrednio nie stosując wielomianów Legendre'a. Rachunki są przy tym wprowadzić znacznie zawilsze i dłuższe, ale za to nie wprowadza się do nich żadnych dodatkowych, upraszczających założeń. Wyciągi, z ułożonych w tym celu tablicy obliczeń, podano na końcu pracy.

Przypadek a. Podstawiając $q' = 0$ ($\kappa = \infty$) do równania (108) znajdujemy

$$\Delta = \frac{1}{2} (1 - \cos \vartheta) + \frac{1}{2} \sum_{i=1} \left[-k_i \frac{3 - \alpha_i}{\varrho_{\alpha i}} + K_i \frac{\beta_i}{\varrho_{\beta i}} - k_i + K_i \right]. \quad (109)$$

Podstawiamy tu $\cos \vartheta = 0,393407$. Po wykonaniu obliczeń znajdujemy zamiast $\Delta = 0$ (jak być powinno) wartość

$$\Delta = +0,00025.$$

Jak wynika z obliczeń, wartość $\Delta = 0,0190$ odpowiada w tym miejscu różnicy kątów ϑ wynoszącej 1° . Podana w tablicy 1 wartość $\vartheta = 66^\circ 50'$ różni się zatem od wartości prawdziwej (przy zastosowaniu wyżej podanej dokładności obliczenia) o niecałą minutę kątową.

Przypadek b. Podstawowy wzór jest w tym przypadku

$$\begin{aligned} \Delta = 0,25 - \frac{1,5}{\varrho'_M} - 0,75 \cos \vartheta + \frac{1}{2} \sum_{i=1}^4 \left(-1,5 k_i \frac{3 - \alpha_i}{\varrho_{\alpha i}} - K_i \frac{3 - \alpha'_i}{\varrho'_{\alpha i}} + \right. \\ \left. + 1,5 K_i \frac{\beta_i}{\varrho_{\beta i}} + k_i \frac{\beta'_i}{\varrho'_{\beta i}} + \frac{1}{2} K_i - \frac{1}{2} k_i \right). \end{aligned} \quad (110)$$

Podstawiając tu $\cos \vartheta = -0,555328$ po dokonaniu obliczeń znajdujemy

$$\Delta = -0,0099.$$

W okolicy tego punktu różnica kąta ϑ o 1° wywołuje zmianę Δ o $0,0528$. Błąd podanej w tablicy 1 wartości $\vartheta = 123^\circ 44'$ wynosi tu zatem $+11'$.

Przypadek c. Mamy $\cos \vartheta = -1$ i podstawowy wzór jest zatem

$$\Delta = 0,125 - \frac{1}{2} \sum_{i=1}^4 \left[K_i \left(\frac{1}{4 - \alpha'_i} - \frac{1}{1 + \beta_i} \right) + k_i \left(\frac{1}{1 + \beta'_i} - \frac{1}{4 - \alpha_i} \right) \right], \quad (111)$$

skąd wynika

$$\Delta = 0,0002.$$

to znaczy, że odchylenie rezultatu obliczenia od właściwej wartości kąta jest minimalnie małe.

Z obliczeń wynika, że liczby podane w tablicy 1 dla $\delta=3$ nieznacznie odbiegają od wartości odnośnych kątów, obliczonych odmienną metodą i z większą dokładnością. Po uwzględnieniu obliczonych korekcji pierwszy wiersz tablicy 1 będzie

$\alpha=1$	1,5	2	3	4	5	∞
$\theta=180^\circ$	$123^\circ 55'$	$109^\circ 53'$	$96^\circ 27'$	$89^\circ 22'$	$85^\circ 5'$	$66^\circ 49'$

W przypadku $\delta=4$ wartości korekcji są jeszcze znacznie mniejsze; przy $\delta>4$, w granicach podanej dokładności korekcje równają się zeru.

Wyciąg z obliczenia w przypadku a. rozdz. 10/2

i	a_i	q_{ai}	k_i	$\frac{3-a_i}{q_{ai}} k_i$	β_i	$q_{\beta i}$	K_i	$\frac{\beta_i}{q_{\beta i}} K_i$
1	0,33333	2,4521	0,33333	0,36250	0,3750	0,91955	0,12500	0,05098
2	0,38095	2,4080	0,04762	0,05179	0,38182	0,91944	0,01818	0,00755
3	0,38194	2,4071	0,00694	0,00755	0,38196	0,91943	0,00265	0,00110
4	0,38197	2,4071	0,00118	0,00128	0,38197	0,91943	0,00046	0,0019
				0,38907	0,42312		0,14629	0,05982

Wyciąg z obliczenia w przypadku b. rozdz. 10/2

i	q_{ai}	$1,5 \frac{3-a_i}{q_{ai}}$	$q'_{\beta i}$	$\frac{\beta'_i}{q'_{\beta i}}$	H'_i	$H'_i k_i$	Uwaga
1	3,3276	1,20207	1,2171	0,27387	0,92820	0,30940	$H'_i = 1,5 \frac{3-a_i}{q_{ai}} - \frac{\beta'_i}{q'_{\beta i}}$
2	3,2815	1,19719	1,2523	0,30420	0,89299	0,04252	
3	3,2805	1,19709	1,2530	0,30484	0,89225	0,00619	
4	3,2805	1,19707	1,2530	0,30484	0,89223	0,00105	
0,35916							

i	q'_{ai}	$1,5 \frac{3-a'_i}{q'_{ai}}$	$q_{\beta i}$	$1,5 \frac{\beta_i}{q_{\beta i}}$	H_i	$H_i K_i$	Uwaga
1	3,2873	0,79853	1,2478	0,45081	0,34772	0,04346	$H_i = \frac{3-a'_i}{q'_{ai}} - 1,5 \frac{\beta_i}{q_{\beta i}}$
2	3,2807	0,79806	1,2529	0,45712	0,34094	0,00620	
3	3,2805	0,79806	1,2530	0,45726	0,34082	0,00090	
4	3,2805	0,79806	1,2531	0,45723	0,34083	0,00016	

0,05072

Wyciąg z obliczenia w przypadku c. rozdz. 10/2

i	$\frac{K_i}{1+\beta_i}$	$\frac{k_i}{1+\beta'_i}$	$\frac{K_i}{4-a'_i}$	$\frac{k_i}{4-a_i}$
1	0,09091	0,25000	0,03448	0,09091
2	0,01276	0,03448	0,00502	0,01316
3	0,00192	0,00502	0,00073	0,00192
4	0,00033	0,00085	0,00013	0,00033
	0,10592	0,29035	0,04036	0,10632

11. ZAKOŃCZENIE

Krzywe na rys. 10—15 i odnośne tablice mogą służyć nie tylko do obliczania kątów granicznych przy danych potencjałach lub ładunkach, ale również i do stwierdzenia, jak się zmienia kąt ϑ przy zmianie potencjałów, ładunków lub odległości obu kul. Tak więc rozważając tylko najprostsze procesy tego rodzaju ustalamy co następuje:

W układzie dwóch jednakowych kul

- A. 1 Przy zachowaniu stałej odległości między kulami (lub przy zachowaniu niezmienniej wartości parametru δ) i przy zwiększaniu stosunku κ potencjałów obu kul

kąt ϑ maleje.

Jeżeli κ nieograniczenie wzrasta, to kąt ϑ dąży do określonej granicy.

- A. 2 Przy zachowaniu stałej odległości między kulami (lub przy zachowaniu stałej wartości δ) i przy zwiększaniu stosunku χ ładunków obu kul

kąt ϑ maleje.

Jeżeli χ wzrasta nieograniczenie, to kąt ϑ dąży do zera.

- B. 1 Przy zachowaniu stałego stosunku κ potencjałów obu kul i przy zwiększaniu odległości między kulami (lub przy zwiększaniu wartości parametru δ)

kąt ϑ maleje.

Jeżeli δ dąży do ∞ , to ϑ dąży do określonej granicy.

- C. Przy zachowaniu stałego stosunku χ ładunków i przy zwiększaniu odległości między kulami (lub przy zwiększaniu wartości parametru δ)

kąt ϑ rośnie.

Jeżeli δ dąży do ∞ , to ϑ dąży do określonej granicy.

W układzie ładunek punktowy — kula

- D. Przy zachowaniu stałej odległości między ładunkiem punktowym i kulą (lub przy stałym δ) oraz przy zwiększaniu stosunku χ ładunków kąt ϑ maleje.

Jeżeli χ wzrasta nieograniczenie, to kąt ϑ dąży do zera.

- E. Przy zachowaniu stałego stosunku χ ładunków i przy zwiększaniu odległości między ładunkiem punktowym i kulą kąt ϑ rośnie.

Jeżeli odległość wzrasta nieograniczenie, to kąt ϑ dąży do określonej granicy.

Obliczenia kątów granicznych mają w pierwszym rzędzie znaczenie teoretyczne:

1. pogłębiają nasze wiadomości o strukturze pola elektrostatycznego;
2. ułatwiają wyznaczenie graficzne obrazu pola;
3. dają się bezpośrednio przenieść na pole przepływowe prądu elektrycznego.

D o d a t e k (do wzoru 49)

W przypadku $\vartheta = \pi$ dla $r = r'$ mamy

$$F_6 = 1 - \frac{\beta_2}{r} + \frac{\beta_2^2}{r^2} - + \dots = \frac{r}{r + \beta_2},$$

$$F_8 = 1 - \frac{\beta_1}{r} + \frac{\beta_1^2}{r^2} - + \dots = \frac{r}{r + \beta_1},$$

tak że możemy napisać

$$\begin{aligned} A_{\vartheta=\pi} &= \left[\left(\frac{\beta_2 n}{r} - \frac{r m}{D} \right) F_6 + \left(\frac{\beta_1 p}{r} - \frac{r}{D} \right) F_8 - \cos \vartheta \right]_{\vartheta=\pi} \\ &= \frac{D \beta_2 n - r^2 m}{D(r + \beta_2)} + \frac{D \beta_1 p - r^2}{D(r + \beta_1)} + 1. \end{aligned}$$

Przyrównując powyższe wyrażenie do $S - T = 1 + p + n - \frac{r}{D}(m + 1)$ znajdujemy tożsamość

$$\frac{Dn - \beta_2 m}{r + \beta_2} + \frac{Dp - \beta_1}{r + \beta_2} \approx 0.$$

Z zależności (11) i (16) wynika, że $Dp = \beta_1$; poza tym możemy dalej napisać

$$Dn \approx \beta_2 m,$$

гдыж на засады (13) и (16) вынікалобы стад

$$D \frac{p}{2} (R-1) \approx \frac{r^2 D k p}{D^2 k p - r^2} \cdot \frac{R+1}{2k}$$

албо пры застасаваніі (51)

$$\frac{R+1}{R-1} \approx \delta^2 k p - 1. \quad (a)$$

Праву строну рўнўсці (a) можам прадставіць у постаці

$$\delta^2 \frac{\delta^2 - 2}{\delta^2 - 1} - 1 = \delta^2 - 1 - \frac{1}{1 - 1/\delta^2} = \delta^2 - 2 - \frac{1}{\delta^2} - \frac{1}{\delta^4} - \frac{1}{\delta^6} \dots$$

Леву строну рўнўсці (a) можам напісаць на падставе (12)

$$\begin{aligned} \frac{R+1}{R-1} &= \frac{D + \sqrt{D^2 - 4r^2}}{D - \sqrt{D^2 - 4r^2}} = \frac{D^2 - 2r^2 + D\sqrt{D^2 - 4r^2}}{2r^2} = \\ &= \frac{1}{2} \delta^2 - 1 + \frac{\delta^2}{2} \sqrt{1 - \frac{4}{\delta^2}} = \delta^2 - 2 - \frac{1}{\delta^2} - \frac{2}{\delta^4} - \frac{5}{\delta^6} \dots \end{aligned}$$

Відаць стад, же рўзніца міжды обыдвума выражэннямі прыбліжўнўсці рўнўсці (a) выносі

$$\frac{1}{\delta^4} + \frac{4}{\delta^6}.$$

Блўд wzglўдны b , кўтўры попўлнямы пісўаць тў рўнўсць, выносі затем

δ	3	5	10
b	$1,4 \cdot 10^{-3}$	10^{-4}	10^{-6}

Б. КОНОРСКИ

ПРЕДЕЛЬНЫЕ УГЛЫ В ЭЛЕКТРОСТАТИЧЕСКОМ ПОЛЕ ДВУХ ШАРОВ С ЗАРЯДАМИ РАЗНЫХ ЗНАКОВ

Резюме

Предметом обсуждения является электростатическое поле системы двух шаров K и K' на которых находятся заряды $+H$ и $-H'$ причём $H > H'$. Все линии поля исходящие из поверхности шара K' кончаются на поверхности шара K и определяют на ней круг расположенный симметрично относительно оси, соединяющей центры обоих шаров. Прямые проведенные из точек этого круга к центру шара K производят круговой конус с углом при вершине равным 2ϑ . Угол ϑ назван предельным углом; он разделяет поверхность шара K на две части; линии поля исходящие из точек одной из этих частей направляются к шару K' , линии же исходящие из другой — направляются к бесконечности.

В труде указан метод, при помощи которого можно рассчитывать углы ϑ . Метод этот состоит в том, что поле обоих шаров K и K' замещается эквивалентным полем двух последовательностей зарядов, из которых одна расположена внутри поверхности K , другая же внутри поверхности K' . Значения последовательных зарядов в обоих рядах значительно уменьшаются, так что вполне достаточно (как это следует из оценки погрешностей) ограничить до 5 число рассматриваемых зарядов внутри каждого шара. Применяя метод наложения, можно рассчитать результирующий (совокупный) поток ψ_ϑ этих 10 зарядов, проникающий через шарообразную чашу сферической поверхности K , предполагая, что вершинный угол этой чаши равен 2ϑ . Из вышеуказанного следует, что $\psi_\vartheta = H'$; таким образом получается основное уравнение (26), содержащее неизвестный угол ϑ . Учитывая известное значение (27) потенциала V шара K , мы можем упростить уравнение (26) и свести его к виду (32), из которого можно рассчитать угол ϑ .

Для облегчения расчёта — отдельные составные части выражения (32) представлены в виде многочленов Лежандра; значения ϑ мы находим при помощи последовательных проб. Соответственные расчёты были исполнены для следующих случаев:

а) при двух одинаковых шарах для разных значений соотношения их потенциалов (§ 7, табл. I, рис. 10 и 11);

б) при двух одинаковых шарах для разных значений соотношения их зарядов (§ 9, табл. 3, рис. 14 и 15) и

в) в системе точечный заряд-шар, для разных значений соотношения заряда шара к точечному заряду (§ 8, табл. 2, рис. 12 и 13).

Из проведенной оценки погрешностей (§ 10) следует, что как приблизительные вычисления вспомогательных выражений, так и вышеописанный метод вычисления угла ϑ отличаются очень большой точностью.

При помощи представленных семейств кривых мы также можем наблюдать, каким образом изменяется угол ϑ при изменении расстояния между обоими шарами, а также при изменении их зарядов или потенциалов.

Знание угла ϑ углубляет наши сведения об электростатическом поле и позволяет лучше чертить изображения таких полей.

B. KONORSKI

GRENZWINKEL IM ELEKTROSTATISCHEN FELDE ZWEIER KUGELN

Zusammenfassung

Gegenstand der Ausführungen ist das elektrostatische Feld eines Systems zweier Kugeln K , K' , auf welchen sich die Ladungen $+H$ und $-H'$ befinden, wobei $H > H'$ ist. Alle Feldlinien, die von der Kugelfläche K' ausgehen, enden auf der Kugelfläche K und bestimmen auf dieser Fläche einen Kreis, der symmetrisch gegen die beiden Mittelpunkte der Kugeln verbindenden Achse gelegen ist. Die Geraden, die von den Punkten dieses Kreises zum Mittelpunkt der Kugel K geführt werden, bilden einen Kreiskegel mit dem Scheitelwinkel 2ϑ . Der Winkel ϑ wird Grenzwinkel genannt; er teilt die Kugelfläche K in 2 Teile: die Feldlinien, die von den Punkten des einen Teiles ausgehen, münden in Kugel K' ; die Feldlinien, die von den Punkten des anderen Teiles ausgehen, verlaufen in die Unendlichkeit.

In der vorliegenden Arbeit wird eine Methode gezeigt, mittels welcher man imstande ist, die Winkel ϑ zu berechnen. Sie beruht darauf, dass man das Feld des Systems K, K' durch ein gleichwertiges Feld zweier Folgen von Punktquellen ersetzt, von welchen die eine innerhalb der sphärischen Fläche K , die andere — innerhalb der sphärischen Fläche K' gelegen ist. Die Werte der aufeinanderfolgenden Punktladungen beider Folgen sinken rapid, so dass es vollkommen genügt (wie sich aus der Bewertung der entstandenen Fehler ergibt), die Zahl der berücksichtigten Punktladungen auf 5 innerhalb jeder Kugel zu beschränken. Mittels Superposition kann man dann den Kugelausschnitt mit dem Zentralwinkel 2ϑ durchsetzenden Gesamtfluss ψ^ϑ dieser zehn Ladungen berechnen. Auf Grund des vorher Gesagten ist $\psi^\vartheta = H'$; man erhält dadurch die grundlegende Gleichung (26), die den unbekannten Winkel enthält. Wenn man den bekannten Ausdruck (27) für das Potential v der Kugel K berücksichtigt, so kann man die Gleichung (26) vereinfachen und bekommt so schliesslich die Gleichung (32), aus welcher der Winkel ϑ berechnet wird.

Um diese Berechnung zu vereinfachen, werden die einzelnen Summanden des Ausdrucks (32) durch Polynome von Legendre dargestellt; die Werte von ϑ erhält man dann näherungsweise auf dem Wege der aufeinanderfolgenden Probeberechnungen.

Die Berechnungen sind für folgende Spezialfälle durchgeführt worden:

- a) für zwei gleiche Kugeln und für verschiedene Werte des Verhältnisses ihrer Potentiale (§ 7, Tafel 1, Bild 10 u. 11);
- b) für zwei gleiche Kugeln und für verschiedene Werte des Verhältnisses ihrer Ladungen (§ 9, Tafel 3, Bild 14 und 15);
- c) für das System Punktquelle-Kugel und für verschiedene Werte des Verhältnisses der Kugelladung zur Punktladung (§ 8, Tafel 2, Bild 12 und 13).

Aus der durchgeführten Genauigkeitsbewertung der verwendeten Methode folgt, dass die dabei begangenen Fehler ausserordentlich klein sind.

Mit Hilfe der vorgelegten Kurvenscharen sind wir imstande zu bestimmen, wie sich der Grenzwinkel bei Aenderungen des Abstandes beider Kugeln, bezw. bei Aenderung ihrer Potentiale oder Ladungen verändert.

Die Kenntnis des Grenzwinkels gewährt uns einen tieferen Einblick in die Struktur des elektrostatischen Feldes; sie ermöglicht uns auch, die Bilder solcher Felder besser zu zeichnen.



621.3.014.6

S. MINC, Z. FELDBLUM

Zależność między natężeniem prądu ochrony katodowej, a natężeniem prądu błędzącego w rurach metalowych zakopanych w gruncie

Rękopis dostarczono 28. 11. 1955

Wyprowadzono zależność między natężeniem prądu ochrony katodowej, a natężeniem prądu błędzącego w rurach metalowych zakopanych w gruncie. Pozwala ona wyznaczyć minimalne natężenie prądu ochronnego konieczne dla pełnego zahamowania korozji. Uzyskaną zależność sprawdzono laboratoryjnie w modelowym zestawie doświadczalnym i stwierdzono jej słuszność z dokładnością do pewnej empirycznej poprawki, której wprowadzenie jest zrozumiałe ze względu na niemożność praktycznego rozdzielenia korozji gruntowej od korozji prądami błędzącymi. Poprawka ta obejmuje ochronę przed korozją gruntową i odchylenia zestawu od przyjętych założeń przy teoretycznym wyprowadzaniu zależności. Podano ograniczenia, którym podlega w obecnym stanie metoda i zależność.

1. WSTĘP

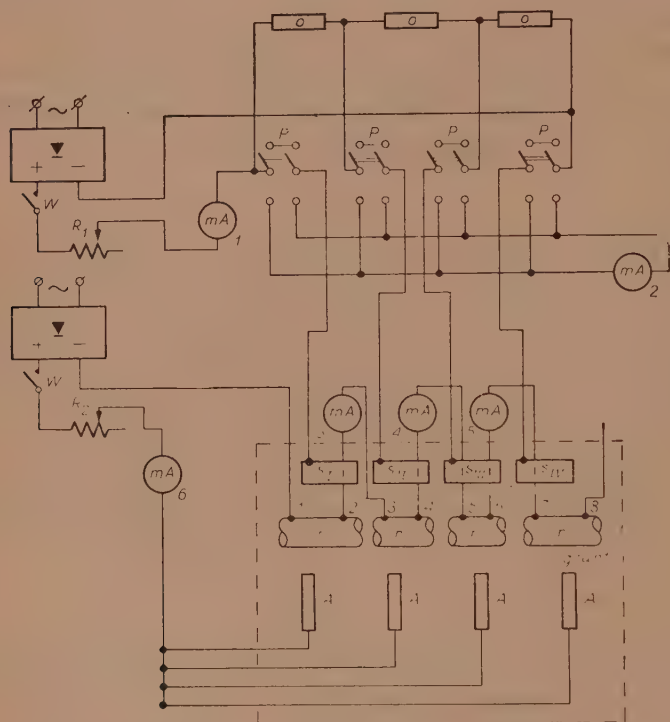
Prace dotyczące korozji powodowanej prądami błędzącymi i ochrony katodowej omawiają zagadnienie to jedynie z punktu widzenia jakościowego i pomiarowego, na przykład [1], [2] [14].

Celem niniejszej pracy było skonstruowanie modelu laboratoryjnego, odpowiadającego przypadkowi korozji rur metalowych zakopanych w gruncie, zachodzącej pod wpływem prądów błędzących (jak i korozji gruntowej), który pozwoliłby teoretycznie wyznaczyć zależność natężenia prądu ochrony katodowej od natężenia prądu błędzącego. Otrzymałą zależność sprawdzono następnie w warunkach laboratoryjnych.

2. BUDOWA I DZIAŁANIE ZESTAWU POMIAROWEGO

Doświadczenia przeprowadzano w drewnianej skrzyni o wymiarach $2 \times 0,5 \times 0,5$ m. Jako grunt służył odpowiednio spreparowany piasek o średnim oporze właściwym rzędu $5 \cdot 10^4 \Omega/\text{cm}$ i wilgotności $4 \div 5\%$. Badana rura żeliwna składała się z czterech części (rurek) o długości 48 cm każda, średnicy zewnętrznej 3,3 cm. Każda rurka posiadała przykręcone

dwie śrubki zaopatrzone w nakrętki, umocowane w odległości 2 cm od obu końców. Służyły one do przyłączania przewodów, w stosunku do których zastosowano numerację 1, 2... 8 (rys. 1). Do każdego pomiaru przygotowywano rurki w następujący sposób: czyszczono je szczotką na toкарке, potem papierem ściernym a następnie wycierano wełnianą szmatką. Rurki umieszczano w gruncie w szeregu jedna za drugą na jednym po-



Rys. 1. Schemat zestawu wytwarzającego prądy błądzące oraz ochrony katodowej rury ($r_I, r_{II}, r_{III}, r_{IV}$)

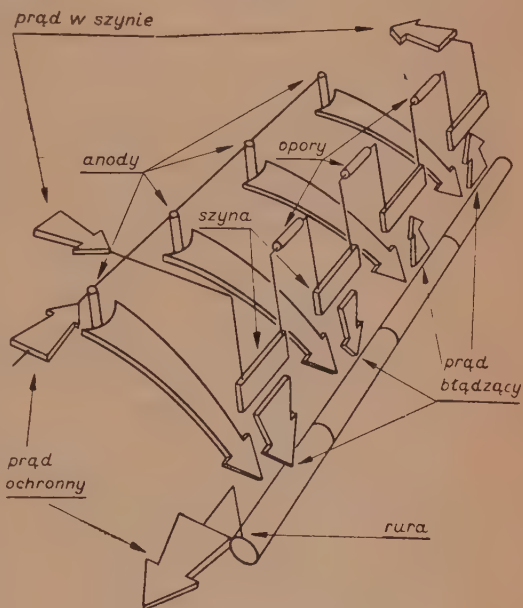
R_1 i R_2 — opory regulacyjne; 20Ω — opory 20Ω ; W — wyłączniki, mA — miliamperomierze; S — szyny; A — anody ochronne; P — przetłączniki.

ziomie. Cały szereg ułożony był symetrycznie, to znaczy w jednakowej odległości od obu ścian bocznych skrzyni. Głębokość zakopania rurek wynosiła 12 cm.

Podstawowym założeniem zestawu elektrycznego było możliwie ściśle zmodelowanie warunków rzeczywistych. Całość zestawu przedstawiono na rys. 1.

Obwód wytwarzający prądy błądzące składał się z prostownika, opornika (R_1) miliamperomierza (1) i szyny. Szyna ta była zbudowana z czterech prostokątnych kawałków blachy stalowej, o wymiarach 22×5 cm

(oznaczonych S_I , S_{II} , S_{III} , S_{IV}); pomiędzy nimi włączono trzy opory po 20 omów. Miało to na celu zapewnienie odpowiedniego spadku napięcia pod wpływem stosunkowo niewielkiego prądu. Spadek ten potrzebny był dla pełniejszego odzwierciedlenia sytuacji panującej w rzeczywistej szynie i spowodowania mierzalnych upływności prądu. Odcinki szyny umieszczano całkowicie w gruncie nad poszczególnymi rurkami, w położeniu do nich równoległym. Do pomiaru prądów upływnych w obwodzie służył miliamperomierz (2), włączany za pomocą przełączników P . Zaznaczyć należy, że prąd upływał właściwie tylko z S_I i S_{II} a do S_{III} i S_{IV} wpływał. Część prądu upływnego płynęła rurą. Był to właśnie prąd błędzący. Dzięki przecięciu rury na cztery części umożliwiono bezpośredni pomiar tego prądu. Mianowicie przewody w położeniu 2—3, 4—5 i 6—7 przyłączano odpowiednio do miliamperomierzy (3, 4 i 5), co dawało możliwość bezpośredniego pomiaru wartości prądu błędzącego w danych punktach. Pozwoliło to wykreślić krzywą natężenia prądu błędzącego na rurze i wyznaczyć maksymalne natężenie prądu błędzącego (położenie 4—5). Natężenie prądu błędzącego wyznaczano także metodą kompensacyjną Radley'a - Richardsa [12]. Pomiarów te dawały wyniki całkowicie zgodne z uzyskanymi metodą bezpośrednią.



Rys. 2. Schemat przepływu prądu błędzącego i ochronnego.

Obwód ochrony katodowej zawierał prostownik, opornik (R_2), miliamperomierz (6) i cztery anody węglowe w kształcie walców o długości 15 cm i średnicy 2 cm (numerowane I, II, III, IV). Anody te umieszczano całkowicie w gruncie, na wysokości połowy każdej rurki w odległości 20 cm od nich. Biegun ujemny prostownika przyłączano do rury w położeniu 1. Tak więc przepływ prądu ochronnego miał kierunek przeciwny do przepływu prądu błędzącego, jak to przedstawiono na rys. 2. Na rysunku tym zilustrowano przepływ prądu błędzącego i ochronnego jedynie w sposób jakościowy, nie przedstawiając rzeczywistej sytuacji powstałej w czasie pełnej ochrony katodowej w stosowanym zestawie, podczas której nie występuje już wypływ prądu błędzącego z rurek 5—6 i 7—8. Prąd

pobierany z gruntu przez ostatnie dwie szyny (S_{III} i S_{IV}) jest wtedy dostarczany przez sąsiadujące z nimi anody ochronne. Całkowity prąd zbierany z rury podczas pełnej ochrony jest sumą prądu błędzającego i ochronnego z pierwszych anod.

3. ANALIZA WSPÓŁDZIAŁANIA PRĄDÓW BŁĄDZĄCYCH I OCHRONNYCH W STOSOWANYM ZESTAWIE

Przy analizie procesu współdziałania prądu błędzającego i ochronnego w stosowanym zestawie przyjęto następujące założenia:

a. stałość przekroju rury, a co za tym idzie, proporcjonalność między odległością dowolnych dwóch przekrojów rury, a opornością części rury między nimi zawartej,

b. liniową zależność między spadkiem napięcia w szynie, a odległością od końca szyny,

c. liniową zależność gęstości prądu, wpływającego do rury na danym jej odcinku, od odległości do końca rury (prąd efektywnie wpływający jest zgodnie z tym liczony jako ujemny),

d. rozmieszczenie ciągle anod stosowanych do ochrony katodowej, co oznacza, że traktuje się cały rząd anod jako jedną anodę, wzdłuż której występujący opór można zaniedbać,

e. stałość oporu właściwego gruntu w całej jego masie,

f. przeciwny kierunek przepływu prądu błędzającego i ochronnego.

W oparciu o charakterystykę stosowanego zestawu i przytoczone założenia, postaramy się znaleźć zależność między natężeniem prądu błędzającego płynącego rurą, a natężeniem prądu ochronnego potrzebnym do uzyskania pełnej ochrony.

Zgodnie z założeniem c. mamy dla gęstości prądu $i_b(x)$, wpływającego do rury, równanie

$$i_b(x) = ax + b, \quad (1)$$

gdzie x jest odległością od początku rury.

Wartości stałych a i b wyznaczamy z podstawowego warunku

$$\int_0^l i_b(x) dx = \int_0^l (ax + b) dx = 0,$$

gdzie l jest całkowitą długością rury; stąd

$$\frac{a}{2} l^2 + bl = 0, \quad \text{skąd} \quad a \frac{l}{2} + b = 0,$$

ponadto $i_b(0) = a \cdot 0 + b = b$. Zatem (1) można napisać w sposób następujący:

$$i_b(x) = i_b(0) \left(1 - \frac{2}{l} x \right). \quad (2)$$

Natężenie prądu (błądzącego) płynącego w rurze $I_b(x)$ można wyrazić wzorem

$$I_b(x) = \int_0^x i_b(x) dx. \quad (3)$$

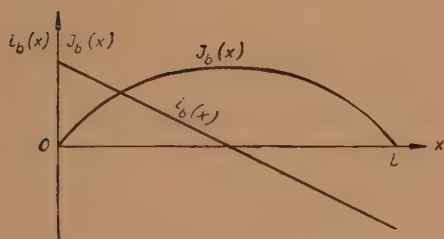
Podstawiając do (3) $i_b(x)$ z (2) otrzymujemy

$$I_b(x) = i_b(0) \int_0^x \left(1 - \frac{2}{l} x\right) dx = i_b(0) \left(x - \frac{x^2}{l}\right). \quad (4)$$

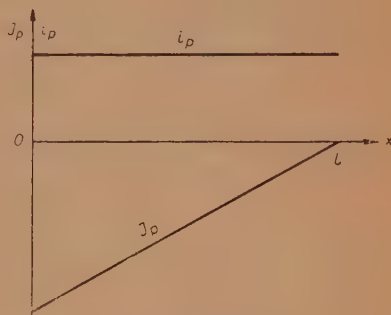
Przebieg zależności $i_b = i_b(x)$ oraz $I_b = I_b(x)$ przedstawiono na rys. 3.

Zależność natężenia prądu w rurze podczas ochrony obliczamy w następujący sposób.

Oznaczmy przez $i_p(x)$ gęstość prądu ochronnego, stałą na całej długości rury, przez $I_p(x)$ — tę część natężenia prądu w rurze, która wynika z prze-



Rys. 3. Wykres zależności $i_b = i_b(x)$ oraz $I_b = I_b(x)$.



Rys. 4. Wykres zależności $i_p = i_p(x)$ oraz $I_p = I_p(x)$.

plywu prądu ochronnego. Wskutek tego, że prąd ochronny zbierany jest w punkcie o współrzędnej $x=0$, jego kierunek będzie przeciwny do $I_b(x)$, co pokazano na rys. 4 przez przypisanie prądowi $I_p(x)$ wartości ujemnych.

Całkowity prąd płynący w rurze w czasie ochrony jest

$$I_c(x) = I_b(x) + I_p(x). \quad (5)$$

$I_p(x)$ wyrażamy jako funkcję $i_p(x)$

$$I_p(x) = i_p(x)(x-l), \quad (6)$$

a podstawiając do (5) wartości z (4) i (6) otrzymujemy

$$I_c(x) = i_b(0) \left(x - \frac{x^2}{l}\right) + (x-l)i_p.$$

Po przekształceniu mamy

$$I_c(x) = -i_b(x) \cdot \frac{x^2}{l} + (i_b(0) + i_p)x - l \cdot i_p. \quad (7)$$

Szukamy takiego warunku na i_p , aby dla $x = \frac{l}{2}$ wartość $I_c(x) = 0$, to znaczy iż powinna być spełniona zależność

$$i(0) \frac{x^2}{l} - (i_b(0) + i_p)x + l \cdot i_p = 0.$$

Stąd po przekształceniach otrzymujemy dla $x = \frac{l}{2}$

$$i_p = \frac{l}{2} i_b(0). \quad (8)$$

Ponieważ wartości i_p i $i_b(0)$ są trudne do bezpośredniego zmierzenia, zależność (8) należy przedstawić za pomocą wielkości $I_p(x)$ i $I_b(x)$. Całkowity prąd ochronny, płynący rurą a wskazywany przez przyrząd włączony w obwód ochrony katodowej, wyraża się wzorem

$$I_p(0) = l \cdot i_p; \quad (9)$$

z drugiej zaś strony z równania (4) dla $x = \frac{l}{2}$ mamy

$$I_b(x) = I_b \left(\frac{l}{2} \right) = i_b(0) \frac{l}{4}. \quad (10)$$

Podstawiając wartości z (9) i (10) do (8), zależność tę można wyrazić jako

$$I_p^0(0) = 2I_b \left(\frac{l}{2} \right). \quad (11)$$

Równanie (11) wyraża fakt, że dla uzyskania $I_c = 0$ w położeniu $\frac{l}{2}$ należy nałożyć prąd ochronny o dwukrotnie wyższej wartości w stosunku do zmierzonej wartości prądu błędzącego w położeniu $\frac{l}{2}$ przed ochroną.

Analogiczne rozumowanie można oczywiście przeprowadzić dla dowolnej wartości x współrzędnej punktu, w którym — przez nałożenie prądu ochronnego — prąd całkowity $I_c(x)$ płynący w rurze sprowadzony zostaje do zera. Powyżej rozpatrywano przypadek $x = \frac{l}{2}$ ze względu na to, że błędy przy wyliczaniu są dla takiej wartości najmniejsze, a ponadto wywody i pomiary — najprostsze. W przypadku ogólnym uzyskamy wyrażenie (wyprowadzone w dalszej części)

$$I_p^0(0) = \frac{l \cdot I_b(x)}{l - x}. \quad (11a)$$

Sytuację obrazującą, zgodnie z (11), wynik nałożenia prądu błędzającego i ochronnego w rurze, przedstawia krzywa I (rys. 5). Z wykresu tego widać, że w prawej połowie rury pozostaje pewna część nie podlegająca ochronie. Jest nią mianowicie odcinek $\frac{3}{4}l$ do l , na którym natężenie $I_c(x)$ przyjmuje jeszcze wartości malejące, co spowodowane jest upływaniem prądu z rury do ziemi. W ogólnym przypadku odcinek rury stanowi obszar upływności prądu do gruntu, gdy $\frac{dI_c(x)}{dx} < 0$ (prąd wchodzący do rury liczony jest jako

dodatni). Dla zapewnienia pełności ochrony przed korozyjnym działaniem prądu błędzającego należy tak dobrać gęstość prądu ochronnego, aby stanowiła ona najmniejszą wartość, dla której spełniony zostaje warunek $\frac{dI_c(x)}{dx} \geq 0$ na całej długości rury. Wymaganiom tym odpowiada krzywa II (rys. 5).

W celu wyznaczenia minimalnej gęstości odpowiadającej podanemu warunkowi należy stwierdzić, dla jakiej wartości i_p zachodzi zależność

$$\frac{dI_c(x)}{dx} = -2 \frac{x}{l} i_b(0) + i_p + i_b(0) \geq 0,$$

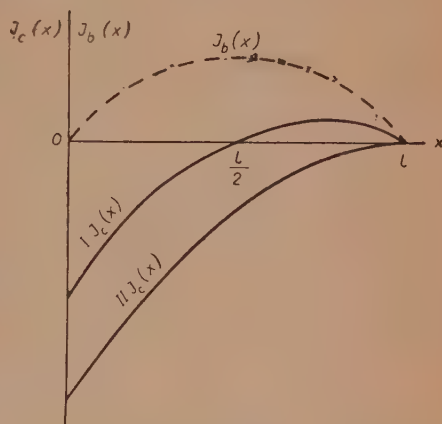
w której prawa strona nierówności pochodzi ze zróżniczkowania wzoru (7). Mamy zatem

$$i_p \geq 2 \frac{x}{l} i_b(0) - i_b(0) \quad (12)$$

dla wszystkich x , skąd dla najmniejszego i_p czyniącego zadość (12), otrzymujemy $i_p = i_b(0)$. Porównując (12) z (8) można wyciągnąć wniosek, że dla uzyskania pełnej ochrony przed korozyjnymi prądami błędzącymi należy zastosować dwukrotnie wyższą wartość prądu ochronnego niż ta, która sprowadza do zera $I_c\left(\frac{l}{2}\right)$. Ponieważ dla uzyskania tego warunku należy, zgodnie

z równaniem (11), nałożyć prąd ochronny $I_p^0(0) = 2I_b\left(\frac{l}{2}\right)$, przeto dla pełnej ochrony powinna być spełniona zależność

$$I_p(0) = 4I_b\left(\frac{l}{2}\right). \quad (13)$$



Rys. 5. Wykres zależności $I_c(x)$
I — krzywa natężenia prądu w rurze
przy $I_c\left(\frac{l}{2}\right) = 0$,
II — krzywa natężenia prądu w rurze
w czasie pełnej ochrony katodowej.

Zależność (13) można przedstawić także w postaci

$$I_p(0) = 2I_p^0. \quad (13a)$$

Wzór (13a) ma tę zaletę, że wartość I_p^0 może być bezpośrednio zmierzona.

3. 1. WSPÓŁDZIAŁANIE I_b i I_p W PRZYPADKU OGÓLNYM

W przypadku ogólnym, zgodnie z równaniem (7), dla $I_c(x) = 0$ mamy

$$\frac{i_b(0)}{l} x^2 - (i_b(0) + i_p)x + l \cdot i_p = 0.$$

Rozwiązaniem tego równania jest

$$x = \frac{i_b(0) + i_p \pm \sqrt{(i_b(0) - i_p)^2}}{2i_b(0)},$$

a pierwiastki tego równania wynoszą

$$x_1 = l, \quad x_2 = \frac{i_p}{i_b(0)} l.$$

Ponieważ zadanie polega na znalezieniu pierwiastka x zawartego w przedziale $0 < x < l$, zatem $x_1 = l$ nie ma sensu, a rozpatrywać należy tylko

$$x = \frac{i_p}{i_b(0)} l$$

lub

$$i_p = \frac{x \cdot i_b(0)}{l}. \quad (8a)$$

Wartości gęstości prądu należy wyrazić wartościami natężenia prądu I_p i I_b . W tym celu do równania (8a) podstawiamy

$$i_p = \frac{l}{I_p(0)} \quad (9)$$

oraz

$$i_b(0) = \frac{l \cdot I_b(x)}{lx - x^2} \quad (4)$$

i uzyskujemy

$$I_p^0(0) = \frac{l \cdot I_b(x)}{l - x}. \quad (11a)$$

Zgodnie z warunkiem (12) mamy $i_b(0) = i_p$. Z porównania tej równości z (8a) wynika, że dla pełnej ochrony katodowej należy prawą stronę równania (11a) pomnożyć przez czynnik $\frac{l}{x}$. Stąd otrzymujemy

$$I_p(0) = \frac{l^2 \cdot I_b(x)}{x(l - x)} \quad (14)$$

lub

$$I_p(0) = \frac{l}{x} I_p^0(0). \quad (14a)$$

3. 2. DOŚWIADCZALNE SPRAWDZENIE ZALEŻNOŚCI (14) I (14a)

Przeprowadzono szereg pomiarów minimalnego natężenia prądu ochrony katodowej powodującego pełne zahamowanie korozji prądami błędzącymi o maksymalnym natężeniu $1 \div 2$ mA. Czas trwania każdego pomiaru wynosił 72 godziny. Stan rury oceniano wizualnie, ponieważ pomiary potencjałów *rura—grunt* nie dały zadowalających, powtarzalnych wyników, wagowy zaś pomiar ubytku masy był niemożliwy ze względów technicznych.

Stwierdzono słuszność uzyskanych teoretycznie wartości I_p z dokładnością do pewnej empirycznej poprawki ΔI_p . Wobec faktu jednoczesnego działania korozji gruntowej i korozji prądami błędzącymi, przy założeniu, że istnieje w przybliżeniu addytywność obu tych form korozji, wartość tę zinterpretowano jako natężenie prądu ochronnego potrzebne dla ochrony przed korozją gruntową. Potwierdził to pomiar w przypadku samej korozji gruntowej. Ponadto stwierdzono, że wartości I_p^0 zmierzone i wyliczone teoretycznie według (14a) nie pokrywają się ściśle ze sobą. Różnica ta

$$\Delta I_p^0 = I_p^0 \text{ emp} - I_p^0 \text{ teor}$$

oznacza, że warunki przeprowadzania pomiarów nie odpowiadały w pełni założeniom postawionym przy wyprowadzeniu badanej zależności. Czynniki, których wynikiem jest ΔI_p^0 , są prawdopodobnie następujące:

niezupełnie liniowy spadek napięcia w szynie,

nieciągłe rozmieszczenie anod ochronnych,

nieznany bliżej charakter współdziałania korozji gruntowej z korozją prądami błędzącymi,

polaryzacja metalu badanej rury.

Przeciętne wartości ΔI_p^0 podano w poniższej tabelicy.

$I_b \text{ max}$ mA	$I_p^0 \text{ emp}$ mA	$I_p^0 \text{ teor.}$ mA	ΔI_p^0 ‰
1,0	2,2 — 2,4	2,0	10 — 20
1,6	3,15 — 3,5	3,2	1,5 — 9,5
2,0	4,4	4,0	10

Wobec konieczności stosowania poprawek na korozję gruntową i *nieidealność* układu, wzory (13) i (13a) należy napisać jako

$$I_p(0) = 4I_b \left(\frac{l}{2} \right) + \sum \Delta I_p \quad (15)$$

lub w przypadku ogólnym

$$I_p(0) = \frac{l^2 \cdot I_b(x)}{x(l-x)} + \sum \Delta I_p. \quad (15a)$$

4. OCENA UZYSKANYCH WYNIKÓW I WNIOSKI

a. Przeprowadzane badania należy ocenić jako uzyskanie metody wyznaczania minimalnego natężenia prądu ochrony katodowej wystarczającego do pełnego zahamowania procesu korozji prądem błądzącym o znanym natężeniu, zmierzonym w określonym położeniu x na rurociągu (ograniczenia metody — rozdz. 4, p. e.). Zależność tę wyraża wzór (15a), gdzie:

l — długość odcinka rury,

$I_p(0)$ — całkowite natężenie prądu ochronnego,

$I_b(x)$ — natężenie prądu błądzącego w położeniu (x) ,

$\sum \Delta I_p$ — empiryczna poprawka na korozję gruntową i lokalne odchylenia wynikające ze współdziałania korozji gruntowej z korozją prądami błądzącymi oraz odchylenia od idealnego rozkładu prądu błądzącego.

b. Dokładnie zbadano doświadczalnie zależność (15a) w przypadku $x = \frac{l}{2}$, dla którego zachodzi zależność (15). Przypadek ten charakteryzuje się dużą prostotą pomiaru i obarczony jest najmniejszym błędem.

c. Wartość

$$\frac{l \cdot I_b(x)}{l-x}$$

ze wzoru (15a) może być wyznaczona bezpośrednim pomiarem. Odpowiada jej taki prąd $I_p(0)$ w położeniu x , dla którego $I_c(x)$ prąd całkowity płynący w rurze) wynosi zero. Mając wyznaczoną doświadczalnie wartość I_p^0 , można ze wzoru (15) dokładniej wyliczyć $I_p(0)$, konieczne dla pełnej ochrony, ponieważ wtedy poprawka $\sum \Delta I_p$ dotyczy jedynie korozji gruntowej.

d. Metoda podana w pracy — po uprzednim sprawdzeniu dla wyższych wartości I_b i dłuższych okresów ochrony — może znaleźć obok laboratoryjnego również zastosowanie praktyczne w warunkach polowych.

e. Stosowalność podanej metody i wzoru (15a) podlega następującym ograniczeniom:

Kierunek prądu ochronnego musi być przeciwny do kierunku przepływu prądu błądzącego (ograniczenie to może mieć duże znaczenie w warunkach polowych).

Metoda dotyczy jedynie przypadku najbardziej drastycznego, a mianowicie równoległego umieszczenia szyny w stosunku do chronionej

rury. Przy zmianie kąta ułożenia względem siebie szyn i rur sytuacja będzie nieco inna.

Układ odzwierciedla sytuację, w której na trasie znajduje się tylko jeden elektrowóz: stąd paraboliczny rozkład prądu upływnego i błędzącego. W przypadku przejazdu kilku elektrowozów na jednej linii — rzeczywisty rozkład prądu błędzącego będzie nałożeniem się kilku parabol.

Wzór (15a) traci znaczenie praktyczne przy niewielkich natężeniach prądu błędzącego a silnej korozji gruntowej, wtedy bowiem wartość $\sum II_p$ jest stosunkowo bardzo duża, a korozja gruntowa jest czynnikiem limitującym cały proces korozji.

f. Wzór (15a) jest podaniem pewnego przepisu na wyliczenie prądu ochronnego przy znajomości natężenia prądu błędzącego, jednakże nie rozwiązuje on na przykład kwestii rozmieszczenia anod ochronnych. Problem taki należy rozpatrywać dodatkowo w oparciu o znajomość charakterystyki danego gruntu. Szereg danych dotyczących tego problemu podano w pracach [1, 2, 6, 12].

WYKAZ LITERATURY

1. Akimow G. W.: *Podstawy nauki o korozji i ochronie metali*. P. W. T. 1952.
2. Cikierman E. J.: *Borba a korrozji podziemnych truboprowodow*, Moskwa 1951.
3. — *Cathodic protection, a symposium, by the National Association of Corrosion Engineers*, 1949.
4. Dziuba W., Ostaszewicz J.: *Prądy błędzące*. W. K. 1955.
5. Evans U. R.: *Metallic corrosion, passivity and protection*. Londyn 1948.
6. — *Korozja metallow* (zbiornik pierewodow statiej), Moskwa 1953.
7. Minc S.: *Bodania nad możliwością ochrony katodowej stalowych ścianek szczelnych przed korozją w basenach portowych Gdańska i Gdyni*. Prace M. I. T. nr 2, 1953.
8. Minc S.: *Przemysł chemiczny*. 29, 556, 1950.
9. Pawlikowski S., Kobylczyk A., Ludera L.: *Zeszyty Naukowe Politechniki Śląskiej — Chemia* nr 1, Kraków 1954.
10. Podolski R.: *Trakcja elektryczna*, t. I, Warszawa.
11. Prituła W. A.: *Zaszcita podziemnych truboprowodow ot wnieszniej korrozji*. Gostoptiechizdat 1948.
12. Radley W. G., Richards C. E.: *J. Inst. Elect. Eng.* nr 8, 1939.
13. Rej B.: *Zarys trakcji elektrycznych*. P. Z. W. S. 1953.
14. — *Zaszcita podziemnych metalicznych truboprowodów i kabellej ot korrozji*. Moskwa 1954.

С. МИНЦ, З. ФЕЛЬДБЛЮМ

ЗАВИСИМОСТЬ МЕЖДУ СИЛОЙ ТОКА КАТОДНОГО ПРЕДОХРАНЕНИЯ
И СИЛОЙ БЛУЖДАЮЩИХ ТОКОВ В ЗАРЫТЫХ
В ЗЕМЛЕ МЕТАЛЛИЧЕСКИХ ТРУБАХ

Резюме

Выведена зависимость между силой тока катодного предохранения и силой блуждающего тока в зарытых в земле металлических трубах. Она позволяет определить наименьшую силу предохранительного тока необходимую для полного предупреждения коррозии. Полученная зависимость была проверена в лаборатории на экспериментальной модели. Её правильность была подтверждена с точностью некоторой эмпирической поправки, введение которой понятно ввиду невозможности практически отличить и разграничить разъедающее действие почвы от коррозии вызываемой блуждающими токами.

Эта поправка содержит защиту перед разъедающим действием почвы и отклонения экспериментальной модели от принятых при теоретическом выводе зависимости предположений. Поданы ограничения, которым подвержены метод и зависимость в настоящем виде.

S. MINC, Z. FELDBLUM

A DEPENDENCE BETWEEN THE CURRENT OF CATHODIC PROTECTION
AND THE STRAY CURRENT IN UNDERGROUND METALLIC PIPES

Summary

A dependence between the current of cathodic protection and the stray current in underground mettalic pipes has been deduced. It permits to find the current required for a complete cathodic protection of the pipes concerned. The exactness of the relation obtained was tested on a laboratory arrangement. In addition an experimental correction has been introduced which is evident. This correction is related exclusively to the corrosive action of the ground and to deviations of the actual experimental arrangement from theoretical pressumptions. The limitations of the applied relation and method are given.

621.396.62.51:62.396.619.13

F. WIŚNIEWSKI

Analiza układu superreakcyjnego z wygaszaniem wywołującym modulację częstotliwości

Rękopis dostarczono 12. 5. 1955

W pracy przedstawiono analizę liniowego układu superreakcyjnego, w którym oprócz zmiennej przewodności występuje również zmienna pojemność wywołująca modulację częstotliwości układu. Otrzymano rozwiązanie o bardziej ogólnym charakterze niż dotychczas znane rozwiązania, ponieważ odnosi się ono zarówno do układów o stałej, jak i zmiennej częstotliwości. Z analizy wynika, że zmiana częstotliwości rezonansowej układu odbiorczego w jego czasowym przedziale czułości powoduje odkształcenie charakterystyki selektywności układu. Podano warunki, w których odkształcenie to będzie niesymetryczne, a wyniki rozwiązań teoretycznych potwierdzono wynikami doświadczalnymi. W końcowej części pracy wykazano możliwość stosowania wzorów otrzymanych z analizy układu liniowego do wyznaczania charakterystyki selektywności układu nieliniowego. W dodatku przedstawiono wyniki doświadczalne wpływu poszczególnych czynników na wielkość odkształcenia charakterystyki selektywności układu.

1. WSTĘP

Charakterystyki selektywności układów superreakcyjnych, obliczone za pomocą ogólnie znanych wzorów [12], są zawsze — przy małej szerokości wstęgi względem częstotliwości rezonansowej układu — krzywymi symetrycznymi. Pomierzone charakterystyki selektywności mają natomiast w wielu przypadkach kształt wyraźnie niesymetryczny. Wartość tej niesymetrii osiąga niekiedy kilkadziesiąt procent całkowitej szerokości charakterystyki i może być różnego znaku, to znaczy w jednych przypadkach charakterystyka jest bardziej rozciągnięta w kierunku mniejszych częstotliwości, w innych — przeciwnie.

W literaturze omawiającej pracę odbiorników superreakcyjnych znajdują się krótkie wzmianki na temat niesymetrycznych krzywych selektywności. Hazeltine [6] i Richman [8] zauważyli, że niesymetria występuje w przypadku znacznej zmiany częstotliwości układu odbiorczego w jego czasowym przedziale czułości. Whitehead [11] podał to samo spostrzeżenie

dodając, że zmiana częstotliwości układu może być wywołana zmianą pojemności międzyelektrodowych lampy w okresie wygaszania oraz składową bierną przewodności zmiennej, którą lampa wnosi do układu odbiorczego.

W celu uzupełnienia powyższych obserwacji przeprowadzono szereg eksperymentów, w wyniku których stwierdzono, że niesymetria charakterystyki selektywności jest ściśle związana ze zmianą częstotliwości układu odbiorczego w jego czasowym przedziale czułości. Wartość niesymetrii zależy od charakterystyki zmian częstotliwości układu odbiorczego, wartości części rzeczywistej współczynnika sprzężenia zwrotnego oraz od kształtu napięcia wygaszającego. Załączony na końcu artykułu dodatek zawiera wyniki i omówienie powyższych doświadczeń.

Opublikowane dotychczas analizy układów superreakcyjnych odnosiły się do układów o stałej częstotliwości rezonansowej i wskutek tego nie mogły uwzględnić przypadku niesymetrycznej charakterystyki selektywności. Analiza przedstawiona w niniejszej pracy obejmuje układy o zmiennej częstotliwości, a otrzymane wyniki mogą być stosowane zarówno do układów o stałej, jak i zmiennej częstotliwości rezonansowej.

2. ANALIZA PRACY UKŁADU SUPERREAKCYJNEGO Z WYGASZANIEM WYWOŁUJĄCYM MODULACJĘ CZĘSTOTLIWOŚCI

2.1. UKŁAD ZASTĘPCZY WZMACNIACZA SUPERREAKCYJNEGO

Jak wiadomo, lampowy układ generacyjny, zawierający oprócz lampy równoległy obwód rezonansowy RLC , można zastąpić układem równoważnym, składającym się z obwodu rezonansowego z dołączoną równolegle przewodnością ujemną wnoszoną przez lampę.

Ponieważ wzmacniacz superreakcyjny stanowi układ generacyjny, w którym warunki generacji są uzależnione od napięcia wygaszającego, może on być zastąpiony obwodem rezonansowym RLC z dołączoną równolegle zmienną przewodnością ujemną wnoszoną przez lampę i uzależnioną od napięcia wygaszającego. Źródło sygnału zewnętrznego można przedstawić w postaci generatora prądowego dołączonego do obwodu.

W ogólnym przypadku przewodność jaką lampa wnosi do układu ma oprócz składowej rzeczywistej również składową urojoną, która powoduje zmianę częstotliwości rezonansowej układu odbiorczego¹. Dla przypadku

¹ Pojęcie przewodności zespolonej układu odnosi się ściśle tylko do napięć i prądów o przebiegu sinusoidalnym. W układach superreakcyjnych zmiany amplitudy drgań w czasie jednego okresu wielkiej częstotliwości są stosunkowo małe i przebieg napięcia w tych układach może być w stosunkowo krótkim przedziale czasu uważany za sinusoidalny. Z tego też względu można z pewnym przybliżeniem mówić o przewodności zespolonej układu superreakcyjnego.

ogólnego, przy założeniu małych zmian częstotliwości w stosunku do częstotliwości rezonansowej, wzmacniacz superreakcyjny może być przedstawiony za pomocą układu zastępczego pokazanego na rys. 1, na którym poszczególne symbole oznaczają:

L — indukcyjność obwodu,

$$g(t, u) = G_0 + G(t, u),$$

gdzie G_0 — stała przewodność rzeczywista obwodu,

$G(t, u)$ — zmienna przewodność rzeczywista wnoszona przez lampę,

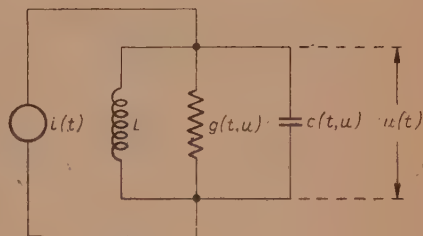
$$c(t, u) = C_0 + C(t, u),$$

gdzie C_0 — stała pojemność obwodu,

$C(t, u)$ — zmienna pojemność wnoszona przez lampę,

$i(t)$ — prąd sygnału płynący przez układ,

$u(t)$ — napięcie na obwodzie.



Rys. 1. Układ zastępczy wzmacniacza superreakcyjnego o zmiennej częstotliwości rezonansowej.

Zagadnienie teoretycznego rozwiązania układu superreakcyjnego sprowadza się do określenia napięcia $u(t)$ występującego na obwodzie, jeżeli są znane pozostałe parametry układu.

2.2. OGÓLNE ROZWIĄZANIE RÓWNIANIA RÓŻNICZKOWEGO DLA UKŁADU O ZMIENNEJ PRZEWODNOŚCI ZESPOLONEJ

Zakładając liniową pracę rozpatrywanego układu, to znaczy, że przewodność zespolona wnoszona przez lampę jest tylko funkcją czasu, a nie zależy od napięcia na obwodzie, można napisać równanie

$$u'c + uc' + gu + \frac{1}{L} \int u dt = i. \quad (1)$$

Po zróżniczkowaniu powyższego wyrażenia względem czasu i podzieleniu stronami przez c otrzymuje się

$$u'' + \left(\frac{g + 2c'}{c} \right) u' + \left(\frac{g' + c''}{c} + \frac{1}{Lc} \right) u = \frac{i'}{c}.$$

Wprowadzając oznaczenia

$$\frac{g + 2c'}{c} = p \quad \text{oraz} \quad \frac{1}{Lc} = \omega^2$$

powyższą równość można przedstawić w postaci

$$u'' + pu' + \left(\frac{g' + c''}{c} + \omega^2 \right) u = \frac{i'}{c}. \quad (2)$$

Podstawienie do równania (2) odpowiednio

$$u = y \exp \left(-\frac{1}{2} \int p dt \right),$$

$$u' = y' \exp \left(-\frac{1}{2} \int p dt \right) - \frac{1}{2} y p \exp \left(-\frac{1}{2} \int p dt \right),$$

$$u'' = y'' \exp \left(-\frac{1}{2} \int p dt \right) - y' p \exp \left(-\frac{1}{2} \int p dt \right) - \frac{1}{2} y p \exp \left(\frac{1}{2} \int p dt \right) +$$

$$\text{daje} \quad + \frac{1}{4} y p^2 \exp \left(-\frac{1}{2} \int p dt \right)$$

$$y'' + \left(-\frac{1}{2} p' - \frac{1}{4} p^2 + \frac{g' + c''}{c} + \omega^2 \right) y = \frac{i'}{c} \exp \left(\frac{1}{2} \int p dt \right). \quad (3)$$

Rozwiązując następnie wyrażenie w nawiasie otrzymuje się

$$\begin{aligned} -\frac{1}{2} p' - \frac{1}{4} p^2 + \frac{g' + c''}{c} + \omega^2 &= -\frac{1}{2} \frac{g'}{c} - \frac{c''}{c} + \frac{1}{2} \frac{g c'}{c^2} + \left(\frac{c'}{c} \right)^2 + \frac{g'}{c} - \\ &- \frac{1}{4} \left(\frac{g}{c} \right)^2 - \frac{g c'}{c^2} - \left(\frac{c'}{c} \right)^2 + \frac{c''}{c} + \omega^2 = \frac{1}{2} \left(\frac{g' c - g c'}{c^2} \right) - \frac{1}{4} \left(\frac{g}{c} \right)^2 + \omega^2 = \\ &= \left(\frac{g}{2c} \right)' - \left(\frac{g}{2c} \right)^2 + \omega^2. \end{aligned}$$

Zakładając z kolei

$$\left| \left(\frac{g}{2c} \right)' - \left(\frac{g}{2c} \right)^2 \right| \ll \omega^2 \quad \text{dla wszystkich} \quad \frac{g}{2c} \quad (4)$$

zamiast równania (3) można rozpatrywać następujące wyrażenie przybliżone

$$y'' + \omega^2 y = \frac{i'}{c} \exp \left(\frac{1}{2} \int p dt \right). \quad (5)$$

Postawione założenia pozwalają zastąpić równanie uproszczone

$$y'' + \omega^2 y = 0$$

wyrażeniem przybliżonym

$$y'' = \{[f(t)]^2 + f'(t)\} y,$$

gdzie

$$[f(t)]^2 = -\omega^2.$$

Wyrażenie to ma rozwiązanie ogólne o postaci

$$y = C_1 \exp \left(j \int \omega dt \right) + C_2 \exp \left(-j \int \omega dt \right).$$

Wprowadzając oznaczenie $\varphi(t) = \int \omega dt$, powyższe wyrażenie można przedstawić jako

$$y = C_1 \exp[j\varphi(t)] + C_2 \exp[-j\varphi(t)] = A \cos \varphi(t) + B \sin \varphi(t).$$

Rozwiązanie szczególne równania pełnego otrzymuje się metodą uzmiennienia stałych [7] stosując wzór

$$y = y_1 \int \frac{Ry_2}{W} dt + y_2 \int \frac{Ry_1}{W} dt.$$

W rozpatrywanym przypadku

$$y_1 = A \cos \varphi(t) \quad y'_1 = -\varphi'(t) A \sin \varphi(t)$$

$$y_2 = B \sin \varphi(t) \quad y'_2 = \varphi'(t) B \cos \varphi(t)$$

$$R = \frac{i'}{c} \exp\left(\frac{1}{2} \int p dt\right)$$

$$W = y_1 y'_2 - y_2 y'_1 = \varphi'(t) \cdot AB.$$

Podstawiając odpowiednie wartości do powyższego wzoru otrzymuje się

$$y = -\cos \varphi(t) \int \frac{i'}{\varphi'(\tau) c} \exp\left(\frac{1}{2} \int p dt\right) \sin \varphi(\tau) d\tau +$$

$$\text{Ponieważ} \quad + \sin \varphi(t) \int \frac{i'}{\varphi'(\tau) c} \exp\left(\frac{1}{2} \int p dt\right) \cos \varphi(\tau) d\tau.$$

$$\sin \varphi(t) \cos \varphi(\tau) - \cos \varphi(t) \sin \varphi(\tau) = \sin [\varphi(t) - \varphi(\tau)],$$

powyższe rozwiązanie można napisać w postaci

$$y = \int \frac{i'}{\varphi'(\tau) c} \exp\left(\frac{1}{2} \int p dt\right) \sin [\varphi(t) - \varphi(\tau)] d\tau.$$

Rozwiązanie ogólne równania pełnego otrzymuje się przez dodanie do powyższego wyrażenia rozwiązania ogólnego równania uproszczonego, co daje

$$y = \int \frac{i'}{\varphi'(\tau) c} \exp\left(\frac{1}{2} \int p dt\right) \sin [\varphi(t) - \varphi(\tau)] d\tau + A \cos \varphi(t) + B \sin \varphi(t).$$

Ponieważ

$$u = y \exp\left(-\frac{1}{2} \int p dt\right)$$

przybliżonym rozwiązaniem równania (1) będzie wyrażenie

$$u = \exp\left(-\frac{1}{2} \int p dt\right) \int \frac{i'}{\varphi'(\tau) c} \exp\left(\frac{1}{2} \int p dt\right) \sin [\varphi(t) - \varphi(\tau)] d\tau +$$

$$+ u_0 \cos [\varphi(t) + \varphi_0] \exp\left(-\frac{1}{2} \int p dt\right),$$

gdzie $u_0 \cos [\varphi(t) + \varphi_0] = A \cos \varphi(t) + B \sin \varphi(t)$.

2.3. ROZWIĄZANIE DLA UKŁADU SUPERREAKCYJNEGO

Właściwa praca układu superreakcyjnego wymaga, aby w części okresu wygaszania odpowiadającej tłumieniu poziom drgań własnych układu spadł do poziomu szumów. Przy spełnieniu powyższego warunku każdy okres wygaszania można rozpatrywać niezależnie od poprzedniego.

Dla poszczególnego okresu wygaszania przyjmuje się następujące charakterystyczne momenty czasowe:

$t = t_1$ — początkowy moment rozpatrywanego okresu wygaszania,

$t = 0$ — moment, w którym przewodność osiąga wartość zerową przechodząc od wartości dodatnich do ujemnych,

$t = t_2$ — moment, w którym przewodność osiąga wartość zerową przechodząc od wartości ujemnych do dodatnich.

Napięcie na obwodzie osiąga wartość maksymalną w chwili $t = t_2$ i zgodnie ze wzorem (6) jest określone równaniem

$$u(t_2) = \exp\left(-\frac{1}{2} \int_{t_1}^{t_2} p dt\right) \int_{t_1}^{t_2} \frac{i'}{\varphi' c} \exp\left(\frac{1}{2} \int_{t_1}^{\tau} p dt\right) \sin [\varphi(t_2) - \varphi(\tau)] d\tau + \\ + u_0 \cos [\varphi(t_2) + \varphi_0] \exp\left(-\frac{1}{2} \int_{t_1}^{t_2} p dt\right), \quad (7)$$

gdzie u_0 oznacza amplitudę napięcia na obwodzie dla $t = t_1$. Dla prądu sygnału o przebiegu sinusoidalnym $i = I \sin \omega_s t$, którego pochodna $i' = \omega_s I \cos \omega_s t$, wzór (7) przybiera postać

$$u(t_2) = \exp\left(-\frac{1}{2} \int_{t_1}^{t_2} p dt\right) \int_{t_1}^{t_2} \frac{I \omega_s}{\varphi' c} \exp\left(\frac{1}{2} \int_{t_1}^{\tau} p dt\right) \cos \omega_s \tau \sin [\varphi(t_2) - \varphi(\tau)] d\tau + \\ + u_0 \cos [\varphi(t_2) + \varphi_0] \exp\left(-\frac{1}{2} \int_{t_1}^{t_2} p dt\right). \quad (8)$$

Przekształcając z kolei wyrażenie $\cos \omega_s \tau \sin [\varphi(t_2) - \varphi(\tau)]$ i pomijając wyrazy zawierające $[\omega_s \tau + \varphi(\tau)]$, ze względu na ich znikomy wpływ na wartość całki, przy założeniu

$$|\omega_s \tau - \varphi(\tau)| \ll |\omega_s \tau + \varphi(\tau)|$$

wzór (8) sprowadza się do wyrażenia

$$u(t_2) = \frac{1}{2} \exp\left(-\frac{1}{2} \int_{t_1}^{t_2} p dt\right) \int_{t_1}^{t_2} \frac{I \omega_s}{\varphi' c} \exp\left(\frac{1}{2} \int_{t_1}^{\tau} p dt\right) \sin [\omega_s \tau - \varphi(\tau) + \varphi(t_2)] d\tau + \\ + u_0 \cos [\varphi(t_2) + \varphi_0] \exp\left(-\frac{1}{2} \int_{t_1}^{t_2} p dt\right), \quad (9)$$

które można przedstawić w następującej postaci wykładniczej

$$u(t_2) = \mathcal{R}_e \left\{ \frac{1}{2j} \exp \left(-\frac{1}{2} \int_{t_1}^{t_2} p dt \right) \int_{t_1}^{t_2} \frac{I \omega_s}{\varphi' c} \exp \left(\frac{1}{2} \int_{t_1}^{\tau} p dt \right) \exp j [\omega_s \tau - \varphi(\tau) + \varphi(t_2)] d\tau + u_0 \exp j [\varphi(t_2) + \varphi_0] \exp \left(-\frac{1}{2} \int_{t_1}^{t_2} p dt \right) \right\}^{\theta_1}. \quad (10)$$

Poszukiwaną wielkością nie jest wartość chwilowa napięcia, lecz jego moduł i w dalszych rozważaniach można pominąć symbol $\mathcal{R}_e$ oraz wyrażenie określające kąt fazowy. Ponadto w normalnie pracujących układach superreakcyjnych moment $t=t_1$ jest wystarczająco odległy od momentu $t=0$, wobec czego drugi składnik po prawej stronie równania (10) nie ma praktycznie wpływu na wartość napięcia w momencie $t=t_2$.

Uwzględniając, że

$$\varphi(t) = \int \omega dt = \int \omega_0 [1 + k(t)] dt = \omega_0 t + \psi(t),$$

zamiast (10) otrzymuje się

$$u(t_2) = \left| \frac{1}{2} \exp \left(-\frac{1}{2} \int_{t_1}^{t_2} p dt \right) \int_{t_1}^{t_2} \frac{I \omega_s}{\omega c} \exp \left(\frac{1}{2} \int_{t_1}^{\tau} p dt \right) \exp j [(\omega_s - \omega_0) \tau - \psi(\tau)] d\tau \right|.$$

Ze względu na dalszą dyskusję celowe jest przedstawienie powyższego wyrażenia w postaci

$$u(t_2) = \left| \frac{1}{2} \exp \left(-\frac{1}{2} \int_0^{t_2} p dt \right) \int_{t_1}^{t_2} \frac{I \omega_s}{\omega c} \exp \left(\frac{1}{2} \int_0^{\tau} p dt \right) \exp [-j\psi(\tau)] \exp [j(\omega_s - \omega_0) \tau] d\tau \right|. \quad (11)$$

Biorąc pod uwagę, że w układach praktycznych są spełnione warunki

$$\begin{aligned} g &\gg c', \\ 1 &\gg k(t), \\ c &\approx C_0, \end{aligned}$$

równość (11) można przedstawić w postaci uproszczonej

$$u(t_2) = \left| \frac{I \omega_s}{2 C_0 \omega_0} \exp \left(\frac{1}{2} \int_0^{t_2} p_0 dt \right) \int_{t_1}^{t_2} \exp \left(\frac{1}{2} \int_0^{\tau} p_0 dt \right) \exp [-j\psi(\tau)] \exp (-j\Omega \tau) d\tau \right|, \quad (12)$$

gdzie

$$p_0 = \frac{g}{C_0} \quad \text{oraz} \quad \Omega = \omega_0 - \omega_s.$$

Otrzymane rozwiązanie posiada bardziej ogólny charakter niż inne dotychczas znane rozwiązania. Odnosi się ono bowiem zarówno do układów o zmiennej, jak i stałej częstotliwości rezonansowej. W przypadku stałej częstotliwości układu odbiorczego wyrażenie (12) sprowadza się do ogólnie znanej postaci

$$u(t_2) = \frac{I}{2C} \frac{\omega_s}{\omega_0} \exp\left(\frac{1}{2} \int_0^{t_2} p dt\right) \int_{t_1}^{t_2} \exp\left(\frac{1}{2} \int_0^{\tau} p dt\right) \exp(-j\Omega\tau) d\tau.$$

Ze wzoru (12) wynika, że maksymalna amplituda napięcia w rozpatrywanym układzie superreakcyjnym jest liniowo zależna od amplitudy sygnału. I dlatego ten rodzaj pracy układu superreakcyjnego jest nazwany rodzajem liniowym.

W czasie od $t=0$ do $t=t_2$ wypadkowa przewodność rzeczywista układu ma wartość ujemną i powoduje wzrost poziomu drgań. Efekt oddziaływania ujemnej przewodności układu na poziom drgań jest wyrażony czynnikiem

$$K_s = \exp\left(-\frac{1}{2} \int_0^{t_2} p dt\right), \quad (13)$$

który określa się jako wzmocnienie superreakcyjne układu. Wzmocnienie to będzie tym większe, im większa będzie na wykresie $p=f(t)$ powierzchnia ograniczenia osią czasu i wartością $p(t)$ w granicach od $t=0$ do $t=t_2$.

Gdyby przewodność układu nie ulegała zmianie, lecz posiadała wartość stałą równą G_0 , to prąd sygnału o częstotliwości równej częstotliwości rezonansowej układu i o amplitudzie I , płynący przez układ, wywołałby na obwodzie napięcie o amplitudzie

$$u_0 = \frac{I}{G_0}.$$

Całkowite wzmocnienie napięciowe układu jest określone jako stosunek maksymalnej wartości amplitudy napięcia na obwodzie do wartości określonej powyżej i wyraża się zależnością

$$K = \left| \frac{G_0}{2C} \frac{\omega_s}{\omega_0} \exp\left(-\frac{1}{2} \int_0^{t_2} p dt\right) \int_{t_1}^{t_2} \exp\left(\frac{1}{2} \int_0^{\tau} p dt\right) \exp[-j\psi(\tau)] \exp(-j\Omega\tau) d\tau \right|.$$

Wyrażenie

$$K_s = \exp\left(-\frac{1}{2} \int_0^{t_2} p dt\right)$$

zostało poprzednio zdefiniowane jako wzmocnienie superreakcyjne. Pozostały czynnik

$$K_r = \left| \frac{G_0}{2C} \frac{\omega_s}{\omega_0} \int_{t_1}^{t_2} \exp \left(\frac{1}{2} \int_0^\tau p dt \right) \exp [-j\psi(\tau)] \exp (-j\Omega\tau) d\tau \right|$$

określa się jako wzmocnienie reakcyjne układu. Dla przypadku $\omega_s = \omega_0$ osiąga ono wartość maksymalną

$$K_{r \max} = \left| \frac{G_0}{2C} \int_{t_1}^{t_2} s(\tau) \exp [-j\psi(\tau)] d\tau \right|, \quad (14)$$

gdzie

$$s(\tau) = \exp \left(\frac{1}{2} \int_0^\tau p dt \right).$$

Czynnik

$$S(\omega) = \left| \frac{\omega_s}{\omega_0} \int_{t_1}^{t_2} s(\tau) \exp [-j\psi(\tau)] \exp (-j\Omega\tau) d\tau \right| \quad (15)$$

przedstawia zależność wzmocnienia napięciowego układu od wzajemnego stosunku częstotliwości sygnału i częstotliwości układu odbiorczego. Czynnik ten określa selektywność układu superreakcyjnego. Przy założeniu $\frac{\omega_s}{\omega_0} \approx 1$, powyższe wyrażenie ma tę samą postać co moduł transformaty

Fouriera dla pojedynczego przebiegu, określonego w granicach od $\tau = t_1$ do $\tau = t_2$ funkcją $f(\tau) = s(\tau) \exp [-j\psi(\tau)]$.

Ze względu na szczególny wpływ wyrażenia $f(\tau) = s(\tau) \exp (-j\psi(\tau))$ na pracę układu superreakcyjnego, wyrażenie to nosi nazwę funkcji czułości i decyduje zarówno o wzmocnieniu reakcyjnym jak i o charakterystyce selektywności układu. Kształt modułu funkcji czułości obrazuje przebieg tłumienia obwodu. Wąski impuls odpowiada szybkim zmianom przewodności układu w otoczeniu momentu wzbudzenia się drgań własnych. Oznacza to, że obwód ma dużą dobroć w stosunkowo krótkim przedziale czasu. W wyniku tego układ ma małe wzmocnienie reakcyjne i jest mało selektywny. Szeroki impuls dowodzi, że obwód był powoli odtłumiany i przez stosunkowo długi okres czasu posiadał dużą dobroć, co w efekcie spowodowało odpowiednio większe wzmocnienie reakcyjne i większą selektywność układu. Natomiast argument funkcji czułości obrazuje zmiany częstotliwości układu, jakie mają miejsce w przedziale czasowym odpowiadającym odtłumianiu obwodu. Dla układu o stałej częstotliwości argument ma wartość zerową i wtedy funkcja czułości jest rzeczywistą funkcją czasu.

2.4. WPŁYW ZMIENNEJ CZĘSTOTLIWOŚCI REZONANSOWEJ UKŁADU SUPERREAKCYJNEGO NA KSZTAŁT CHARAKTERYSTYKI SELEKTYWNOŚCI

Zmiana częstotliwości rezonansowej układu odbiorczego w jego przebiegu czasowym czułości znajduje swój odpowiednik matematyczny w zespolonej wartości funkcji czułości. Poprzednio wykazano, że charakterystyka selektywności układu jest tak uzależniona od funkcji czułości, jak moduł transformaty Fouriera od funkcji, której odpowiada. Na tej podstawie można stwierdzić, że zagadnienie określenia wpływu zmiennej częstotliwości rezonansowej układu superreakcyjnego na kształt charakterystyki selektywności sprowadza się do określenia właściwości modułu transformaty funkcji zespolonej. Właściwości te zostaną określone za pomocą szeregu twierdzeń odnoszących się do teoretycznie możliwych przypadków.

Wprowadza się następujące oznaczenia:

- S_{1p} — część parzysta składowej rzeczywistej transformaty,
- S_{1n} — część nieparzysta składowej rzeczywistej transformaty,
- S_{2p} — część parzysta składowej urojonej transformaty,
- S_{2n} — część nieparzysta składowej urojonej transformaty,
- f_{1p} — część parzysta składowej rzeczywistej funkcji,
- f_{1n} — część nieparzysta składowej rzeczywistej funkcji,
- f_{2p} — część parzysta składowej urojonej funkcji,
- f_{2n} — część nieparzysta składowej urojonej funkcji.

Twierdzenie 1

Jeżeli spośród czterech części transformaty $S = S_{1p} + S_{1n} + j(S_{2p} + S_{2n})$ przynajmniej trzy są różne od zera oraz w przypadku, gdy wszystkie cztery części są różne od zera, jest spełniona nierówność $S_{1p}S_{1n} \neq -S_{2p}S_{2n}$, to moduł transformaty $|S(\omega)|$ jest niesymetryczny względem częstotliwości środkowej, a jej funkcja $f = f_{1p} + f_{1n} + j(f_{2p} + f_{2n})$ ma przynajmniej trzy spośród swych części różne od zera.

Dowód

Stosując znany wzór Fouriera otrzymuje się

$$\begin{aligned}
 2\pi f(t) &= \int_{-\infty}^{+\infty} S(\omega) \exp(j\omega t) d\omega = \\
 &= \int_{-\infty}^{+\infty} S_{1p}(\omega) \exp(j\omega t) d\omega + \int_{-\infty}^{+\infty} S_{1n}(\omega) \exp(j\omega t) d\omega + \\
 &\quad + j \int_{-\infty}^{+\infty} S_{2p}(\omega) \exp(j\omega t) d\omega + j \int_{-\infty}^{+\infty} S_{2n}(\omega) \exp(j\omega t) d\omega = \\
 &= \int_{-\infty}^{+\infty} S_{1p}(\omega) \cos \omega t d\omega + j \int_{-\infty}^{+\infty} S_{1p}(\omega) \sin \omega t d\omega + \int_{-\infty}^{+\infty} S_{1n}(\omega) \cos \omega t d\omega + \\
 &\quad + j \int_{-\infty}^{+\infty} S_{1n}(\omega) \sin \omega t d\omega + j \int_{-\infty}^{+\infty} S_{2p}(\omega) \cos \omega t d\omega - \int_{-\infty}^{+\infty} S_{2p}(\omega) \sin \omega t d\omega + \\
 &\quad + j \int_{-\infty}^{+\infty} S_{2n}(\omega) \cos \omega t d\omega - \int_{-\infty}^{+\infty} S_{2n}(\omega) \sin \omega t d\omega.
 \end{aligned}$$

Z ostatnich ośmiu całek druga, trzecia, szósta i siódma są równe zero [4], a zatem

$$2\pi f(t) = \int_{-\infty}^{+\infty} S_{1p}(\omega) \cos \omega t d\omega - \int_{-\infty}^{+\infty} S_{2n}(\omega) \sin \omega t d\omega + \\ + j \int_{-\infty}^{+\infty} S_{1n}(\omega) \sin \omega t d\omega + j \int_{-\infty}^{+\infty} S_{2p}(\omega) \cos \omega t d\omega.$$

Pierwsza z powyższych całek określa rzeczywistą i parzystą część $f(t)$, druga całka określa rzeczywistą i nieparzystą część $f(t)$, trzecia określa część urojoną nieparzystą, a czwarta — część urojoną parzystą. Jeżeli przynajmniej trzy spośród czterech części transformaty są różne od zera, to funkcja $f(t)$ ma przynajmniej trzy spośród swych części różne od zera.

Dla przypadku, gdy wszystkie cztery części transformaty są różne od zera

$$|S(\omega)|^2 = |S_{1p}(\omega) + S_{1n}(\omega)|^2 + |S_{2p}(\omega) + S_{2n}(\omega)|^2,$$

$$|S(-\omega)|^2 = |S_{1p}(\omega) - S_{1n}(\omega)|^2 + |S_{2p}(\omega) - S_{2n}(\omega)|^2$$

odejmując powyższe wyrażenia stronami otrzymuje się

$$|S(\omega)|^2 - |S(-\omega)|^2 = 4S_{1p}S_{1n} + 4S_{2p}S_{2n}.$$

Aby powyższy wynik był różny od zera, musi być spełniony warunek $S_{1p}S_{1n} \neq -S_{2p}S_{2n}$ i wtedy moduł transformaty jest niesymetryczny.

Twierdzenie 2

Jeżeli transformata $S(\omega)$ rzeczywista ma część parzystą i część nieparzystą, to jej funkcja $f(t)$ jest zespolona i ma część rzeczywistą parzystą, a część urojoną nieparzystą.

D o w ó d

Ponieważ transformata jest rzeczywista zatem

$$S_{2p} = 0 \quad \text{oraz} \quad S_{2n} = 0$$

i funkcja jest określona następującym wzorem

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{1p}(\omega) \cos \omega t d\omega + \frac{j}{2\pi} \int_{-\infty}^{+\infty} S_{1n}(\omega) \sin \omega t d\omega.$$

Pierwszy składnik równania przedstawia rzeczywistą i parzystą część funkcji, drugi — część urojoną i nieparzystą.

Twierdzenie 3

Jeżeli transformata $S(\omega)$ urojona ma część parzystą oraz część nieparzystą, to jej funkcja $f(t)$ jest zespolona i ma część rzeczywistą nieparzystą, a część urojoną parzystą.

D o w ó d

Ponieważ transformata jest urojona przeto

$$S_{1p} = 0 \quad \text{oraz} \quad S_{1n} = 0$$

i wyrażenie określające funkcję przybiera następującą postać:

$$f(t) = \frac{j}{2\pi} \int_{-\infty}^{+\infty} S_{2p}(\omega) \cos \omega t d\omega - \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{2n}(\omega) \sin \omega t d\omega.$$

Powyższe twierdzenia wyczerpują wszystkie możliwe przypadki istnienia niesymetrycznego modułu transformaty i wynika z nich, że koniecznym warunkiem powstania niesymetrycznego modułu transformaty Fouriera jest zespolona wartość funkcji.

W dalszym ciągu zostaną rozpatrzone właściwości modułów transformat, których funkcje mają wartość zespoloną.

Twierdzenie 4

Jeżeli spośród czterech części funkcji $f = f_{1p} + f_{1n} + j(f_{2p} + f_{2n})$ przynajmniej trzy są różne od zera i w przypadku, gdy wszystkie cztery części są różne od zera, jest spełniona nierówność

$$\frac{\int_{-\infty}^{+\infty} f_{1p}(t) \cos \omega t d\omega}{\int_{-\infty}^{+\infty} f_{2p}(t) \cos \omega t d\omega} \neq \frac{\int_{-\infty}^{+\infty} f_{1n}(t) \sin \omega t d\omega}{\int_{-\infty}^{+\infty} f_{2n}(t) \sin \omega t d\omega},$$

to transformata funkcji jest zespolona, a jej moduł $|S(\omega)|$ ma część parzystą i część nieparzystą.

Dowód analogiczny jak w twierdzeniu 1.

Twierdzenie 5

Jeżeli funkcja $f(t)$ jest zespolona i ma tylko części parzyste, to jej transformata (ω) jest zespolona, a jej moduł $|S(\omega)|$ jest symetryczny względem częstotliwości środkowej.

D o w ó d

Ponieważ części nieparzyste są równe zeru więc

$$S(\omega) = \int_{-\infty}^{+\infty} f_{1p}(t) \cos \omega t dt + j \int_{-\infty}^{+\infty} f_{2p}(t) \cos \omega t d\omega$$

i moduł transformaty jako pierwiastek z sumy kwadratów dwóch funkcji parzystych, jest symetryczny względem częstotliwości środkowej.

Twierdzenie 6

Jeżeli funkcja $f(t)$ jest zespolona i ma tylko części nieparzyste, to jej transformata $S(\omega)$ jest zespolona, a jej moduł $|S(\omega)|$ jest symetryczny względem częstotliwości środkowej.

Dowód analogiczny jak w twierdzeniu 5.

Twierdzenie 7

Jeżeli funkcja zespolona $f(t)$ ma tylko dwie części, z których jedna jest parzysta, a druga nieparzysta, to jej transformata $S(\omega)$ jest rzeczywista lub urojona i niesymetryczna względem częstotliwości środkowej.

D o w ó d

Przypadek 1

$$f_{1p} = 0, \quad f_{1n} = 0$$

stąd

$$S(\omega) = -j \int_{-\infty}^{+\infty} f_{1n}(t) \sin \omega t dt + j \int_{-\infty}^{+\infty} f_{2p}(t) \cos \omega t dt$$

Transformata ma część parzystą i nieparzystą, czyli jest niesymetryczna względem częstotliwości środkowej.

Przypadek 2

$$f_{1n}=0, \quad f_{2p}=0,$$

stad

$$S(\omega) = \int_{-\infty}^{+\infty} f_{1p}(t) \cos \omega t dt + \int_{-\infty}^{+\infty} f_{2n}(t) \sin \omega t dt.$$

Transformata ma część parzystą i nieparzystą i jest niesymetryczna względem częstotliwości środkowej.

W rzeczywistych układach superreakcyjnych mogą mieć miejsce następujące przypadki:

Przypadek 1

Zmiany przewodności i częstotliwości układu odbiorczego mają przebiegi symetryczne względem momentu powstawania drgań własnych układu ($t=0$), co daje składową rzeczywistą i składową urojoną funkcji czułości w postaci funkcji parzystych.

Charakterystyka selektywności takiego układu będzie symetryczna, lecz możliwe jest jej odkształcenie i poszerzenie w stosunku do charakterystyki selektywności odpowiadającej układowi o stałą częstotliwości.

Przypadek 2

Funkcja czułości ma wartość zespoloną, a jej składowa rzeczywista lub urojona, albo też obie, mają części parzyste i nieparzyste równocześnie.

Przypadek ten odpowiada

- a. niesymetrycznemu przebiegowi zmian przewodności rzeczywistej przy symetrycznym przebiegu zmian częstotliwości względem $t=0$;
- b. niesymetrycznemu przebiegowi zmian częstotliwości przy symetrycznym przebiegu zmian przewodności rzeczywistej względem $t=0$;
- c. niesymetrycznym przebiegom zmian przewodności rzeczywistej i częstotliwości względem $t=0$.

Charakterystyka selektywności takiego układu będzie niesymetryczna z wyjątkiem przypadków szczególnych, określonych w założeniach twierdzenia 4. Możliwe jest również dodatkowe jej odkształcenie i poszerzenie w stosunku do charakterystyki odpowiadającej układowi o stałą częstotliwości.

Przypadki omawiane w twierdzeniach 2, 3, 6 i 7 nie występują w układach superreakcyjnych, ponieważ wymagają przyrostów przewodności i częstotliwości o różnych znakach po obu stronach momentu $t=0$, co jest niezgodne z normalnymi warunkami pracy układu.

3. PORÓWNANIE WYNIKÓW OBLICZENIOWYCH Z WYNIKAMI DOŚWIADCZALNYMI DLA CHARAKTERYSTYK SELEKTYWNOŚCI UKŁADÓW O ZMIENNEJ CZĘSTOTLIWOŚCI REZONANSOWEJ

3.1. ŹRÓDŁA ZMIAN CZĘSTOTLIWOŚCI REZONANSOWEJ UKŁADU SUPERREAKCYJNEGO

Zasadniczo można wyodrębnić dwa czynniki powodujące zmianę częstotliwości odbiornika superreakcyjnego: pierwszym z nich jest zmiana pojemności międzyelektrodowych, drugim — reaktancyjne działanie lampy.

Wskutek napięcia wygaszającego zmienia się wartość chwilowa prądu płynącego przez lampę, wobec czego następuje zmiana pojemności międzyelektrodowych w takt zmian ładunku przestrzennego. Procentowa zmiana częstotliwości układu wywołana tą samą zmianą pojemności międzyelektrodowych będzie wzrastać w kierunku większych częstotliwości. Wynika to stąd, że ze wzrostem częstotliwości maleje pojemność obwodu i te same zmiany pojemności międzyelektrodowych spowodują procentowo większe zmiany pojemności wypadkowej.

Drugim czynnikiem powodującym zmianę częstotliwości jest reaktancyjne działanie lampy. Dla triody, w układzie z dodatnim sprzężeniem zwrotnym, słuszny jest wzór [5].

$$\frac{du_a}{di_a} = \frac{\varrho_a}{1 - mK_a} = \frac{1}{\frac{1}{\varrho_a} - mS_a},$$

gdzie

du_a — przyrost potencjału anody,

di_a — przyrost prądu anodowego,

ϱ_a — oporność wewnętrzna lampy,

K_a — współczynnik amplifikacji lampy,

S_a — nachylenie charakterystyki prądu anodowego,

m — współczynnik sprzężenia zwrotnego.

Na ogół współczynnik sprzężenia zwrotnego nie jest wartością rzeczywistą i ma postać ogólną

$$m = m_1 + jm_2.$$

Wstawiając to wyrażenie do poprzedniego wzoru otrzymuje się

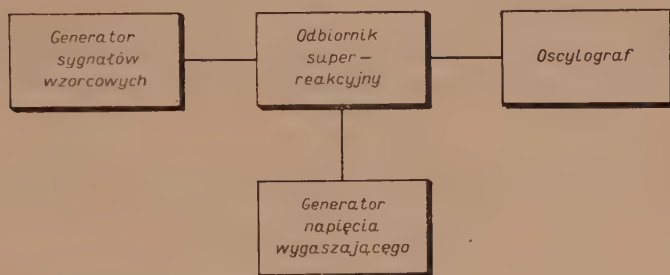
$$\frac{du_a}{di_a} = \frac{1}{\frac{1}{\varrho_a} - m_1 S_a - jm_2 S_a}.$$

Równanie powyższe przedstawia trzy oporności połączone równolegle.

Dwie z nich są rzeczywiste, a trzecia — urojona. Wyraz $-\frac{1}{jm_a S_a}$ przedstawia oporność urojoną dołączoną równolegle do wewnętrznej oporności lampy. Reaktancja ta zmienia swą wartość ze zmianą S_a powodując zmianę częstotliwości obwodu rezonansowego dołączonego do lampy.

3.2. METODY POMIAROWE

3.2.1. *Układ pomiarowy.* W celu doświadczalnego potwierdzenia słuszności wyników przedstawionej analizy, wykonano obliczenia i pomiary dla dwóch układów doświadczalnych pracujących na częstotliwości około 500 kHz. Ze względu na dużą pojemność zewnętrzną, użytą w tych układach, pominięto wpływ, jaki wywiera zmiana pojemności międzyelektrodowych lampy na częstotliwość układu. W obliczeniach uwzględniono



Rys. 2. Schemat blokowy układu pomiarowego.

tylko zmienną przewodność zespoloną, którą lampa wnosi do układu wskutek zespolonej wartości współczynnika sprzężenia zwrotnego.

W układzie pomiarowym, którego schemat blokowy pokazano na rys. 2, generator sygnałów wzorcowych jest źródłem sygnałów w. cz. o znanym poziomie i częstotliwości. Wskaźnikiem poziomu drgań w. cz. w układzie superreakcyjnym jest przeskalowany oscylograf, dołączony na stałe do układu między anodą i katodą lampy superreakcyjnej. Specjalny generator dostarcza napięcia wygaszającego o znanym kształcie, wartości i częstotliwości.

3.2.2. *Pomiar charakterystyki selektywności.* Charakterystykę selektywności odbiornika superreakcyjnego pracującego liniowo określa zależność istniejąca między wartością a częstotliwością sygnału wejściowego, dającego stały poziom drgań w. cz. na wyjściu odbiornika.

W układzie pomiarowym, jak na rys. 2, przy określonych napięciach zasilających odbiornik przykładą się na wejście układu sygnał o częstotliwości dostrojenia odbiornika i o takiej wartości, aby obraz drgań oglą-

dany na oscylografie był dostatecznie duży oraz aby drgania nie dochodziły do poziomu ustalonego, lecz liniowo zmieniały swój poziom ze zmianą wartości sygnału wejściowego. Następnie zmienia się częstotliwość przyłożonego sygnału, zwiększając jednocześnie jego poziom, aby utrzymać stałą wysokość obrazu drgań. Uzyskana w ten sposób charakterystyka określa selektywność badanego odbiornika.

3.3. METODY OBLICZENIOWE

3.3.1. *Funkcja przewodności zespolonej układu superreakcyjnego.* Dla zadanych elementów układu oblicza się stosunek napięcia występującego między siatką a katodą do napięcia występującego między anodą a katodą oraz przesunięcie fazowe między tymi napięciami. Przedstawiając ten stosunek w postaci

$$\hat{m} = m \exp(j\varphi) = m_1 + jm_2$$

wyznacza się część rzeczywistą i część urojoną współczynnika sprzężenia zwrotnego. Następnie, za pomocą charakterystyk lampy, określa się [12] przebieg funkcji zmiennej przewodności rzeczywistej wnoszonej przez lampę

$$G(t) = -m_1 S_a + G_a$$

oraz przebieg funkcji zmiennej przewodności urojonej

$$jB(t) = -jm_2 S_a.$$

3.3.2. *Charakterystyka selektywności.* Charakterystyka selektywności układu superreakcyjnego o zmiennej częstotliwości jest określana za pomocą wyrażenia (15) o postaci

$$S(\omega) = \left| \int_{-\infty}^{+\infty} s(\tau) \exp[-j\psi(\tau)] \exp(-j\Omega\tau) d\tau \right|.$$

Powyższe wyrażenie ma postać modułu transformaty Fouriera przebiegu określonego za pomocą funkcji zmiennej zespolonej.

Przed przystąpieniem do wykonywania właściwych działań — zgodnie ze wzorem Fouriera — należy wyznaczyć przebieg funkcji czułości. Moduł tej funkcji

$$s(\tau) = \exp\left(\frac{1}{2} \int_0^{\tau} p_0 dt\right)$$

oblicza się metodą wykreślną [12]. W celu wyznaczenia przebiegu argumentu $\psi(\tau) = \int_0^{\tau} \Delta\omega dt$ należy powiązać wartości parametrów układu z argumentem funkcji czułości.

Jak wiadomo, częstotliwość chwilowa układu jest określona wzorem

$$\omega = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{LC_0} \sqrt{1 + \frac{\Delta C}{C_0}}} = \omega_0 \left(1 - \frac{\Delta C}{2C_0}\right) = \omega_0 + \Delta\omega,$$

gdzie $\Delta C = C - C_0$ oraz $\Delta\omega = \omega - \omega_0$

skąd przyrost częstotliwości

$$\Delta\omega = -\frac{\Delta C}{2C_0} \omega_0.$$

Lampa wnosi do układu przewodność urojoną, której wartość

$$\Delta B = \omega \Delta C \approx \omega_0 \Delta C.$$

Podstawiając powyższe wyrażenie do wzoru na przyrost częstotliwości otrzymuje się

$$\Delta\omega = -\frac{\Delta C}{2C_0} \omega_0 = -\frac{\Delta B}{2C_0}$$

i zależność argumentu funkcji czułości od parametrów układu można przedstawić jako

$$\psi(\tau) = \int_0^\tau \Delta\omega dt = -\frac{1}{2C_0} \int_0^\tau \Delta B dt = -\frac{m_0}{2C_0} \int_0^\tau \Delta S_a dt.$$

Ponieważ funkcja czułości może mieć przebieg trudny do przedstawienia w sposób analityczny, wyznaczenie widma w takim przypadku odbywa się zwykle za pomocą metody analityczno-wykreślnej. W tym celu należy analizowany przebieg rozłożyć na część rzeczywistą i część urojoną. Opierając się na liniowości przekształcenia Fouriera autor rozkładał następnie każdą z tych części na możliwie najmniejszą ilość figur prostych jak trójkąt i prostokąt, dla nich wyznaczał transformaty i odpowiednio je dodawał.

3.4. WYNIKI POMIARÓW I OBLICZEŃ

Przykład 1

Dane układu

Pomiary i obliczenia wykonano dla układu superreakcyjnego o następujących danych:

$$\begin{aligned} U_{a0} &= 175 \text{ V}, & U_{wyg} &= 4,5 \text{ V}_{sk}, \text{ (sinusoidalne)} \\ U_{s0} &= -8,0 \text{ V}, & f_{wyg} &= 1,5 \text{ kHz}, \\ U_{wzb} &= -5,4 \text{ V}, & f_0 &= 525 \text{ kHz}, \end{aligned}$$

gdzie

U_{a0} — stałe napięcie anodowe,

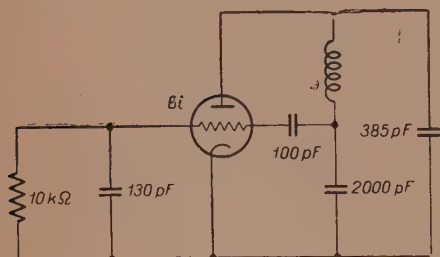
U_{s0} — stałe napięcie siatkowe,

U_{wzb} — napięcie siatkowe przy którym wzbudzają się drgania własne układu,

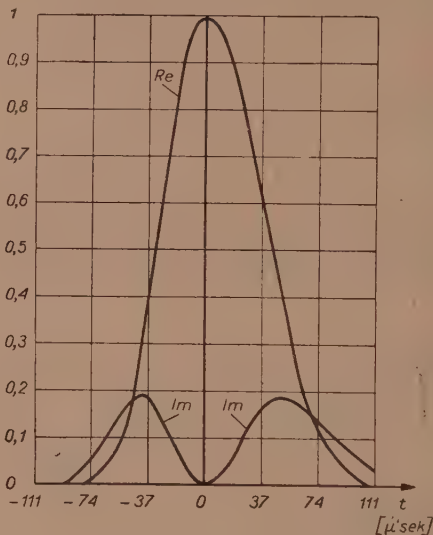
U_{wyg} — napięcie wygaszające,

f_{wyg} — częstotliwość napięcia wygaszającego,

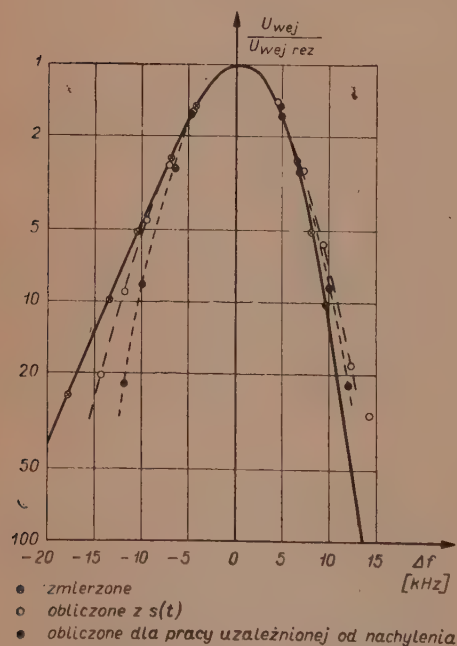
f_0 — częstotliwość rezonansowa układu.



Rys. 3. Uproszczony schemat układu superreakcyjnego.

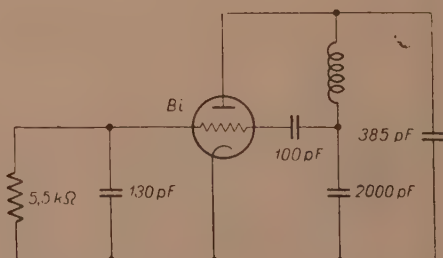


Rys. 4. Funkcja czułości.



Rys. 5. Charakterystyka selektywności układu.

Uproszczony schemat omawianego układu jest przedstawiony na rys. 3, a wyniki obliczeń i pomiarów są przedstawione na rys. 4 i 5.



Rys. 6. Uproszczony schemat układu superreakcyjnego.

Przykład 2

Dane układu

Pomiary i obliczenia wykonano dla układu superreakcyjnego o następujących danych:

$$U_{a0} = 175 \text{ V,}$$

$$U_{wyg} = \text{[symbol trapezoidal pulse]}$$

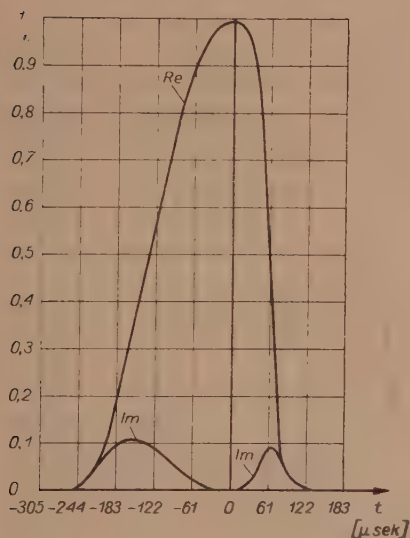
$$U_{s0} = -6,8 \text{ V,}$$

$$f_{wyg} = 1820 \text{ Hz,}$$

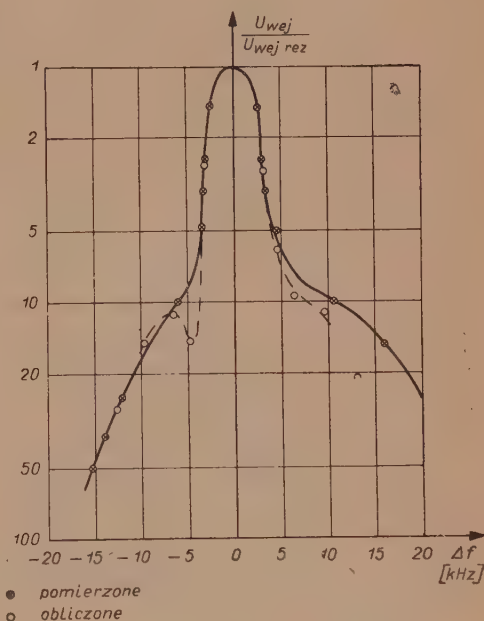
$$U_{wzb} = -5,6 \text{ V,}$$

$$f_0 = 525 \text{ kHz,}$$

Uproszczony schemat omawianego układu jest przedstawiony na rys. 6, natomiast wyniki obliczeń łącznie z wynikami pomiarów są przedstawione na rys. 7 i 8.



Rys. 7. Funkcja czułości.



Rys. 8. Charakterystyka selektywności układu.

3.5. WNIOSKI

Przykład pierwszy dotyczy układu superreakcyjnego z wygaszaniem sinusoidalnym, dla którego bez uwzględnienia zmiany częstotliwości rezonansowej stosuje się wzory dla pracy uzależnionej od nachylenia. Rys. 5 przedstawia trzy charakterystyki selektywności: zmierzoną, obliczoną według wzoru dla pracy uzależnionej od nachylenia i obliczoną metodą widmową, z uwzględnieniem zmiany częstotliwości układu odbiorczego.

Najbardziej wąska jest krzywa obliczona według wzoru dla pracy uzależnionej od nachylenia. Krzywa obliczona metodą widmową pokrywa się z krzywą zdjętą na drodze eksperymentalnej do 5-krotnego spadku wzmocnienia, a dalej wykazuje pewną rozbieżność, choć charakter przebiegu obu krzywych pozostaje ten sam. Występująca rozbieżność jest prawdopodobnie spowodowana nieścisłością wykreślonej metody obliczania widma. Z przykładu widać, że zmiana częstotliwości układu odbiorczego powoduje nie tylko niesymetrię charakterystyki, lecz również jej poszerzenie, czyli zmniejszenie selektywności układu.

Przykład drugi dotyczy układu superreakcyjnego, w którym zastosowano napięcie wygaszające o specjalnym kształcie w celu zwiększenia selektywności układu przy zachowaniu stałej częstotliwości wygaszania i stałej wartości wzmocnienia.

Na rys. 8 przedstawiono dwie charakterystyki selektywności: zmierzoną i obliczoną. Podobnie jak w przykładzie pierwszym obie charakterystyki pokrywają się do pięciokrotnego zmniejszenia wzmocnienia. Dalej występuje pewna rozbieżność między wynikami pomiarów i obliczeń, mimo to charakter obu krzywych jest ten sam. Rozbieżność ta leży w granicach błędu obliczenia transformaty funkcji czułości. W przykładzie drugim niesymetria krzywej selektywności ma przeciwny znak niż w przykładzie pierwszym, przy czym krzywa selektywności ma specyficzny kształt. Jest to spowodowane kształtem napięcia wygaszającego, które oddziałuje na selektywność układu nie tylko na odcinku o najmniejszym nachyleniu, lecz również części o większym nachyleniu wywierają pewien wpływ, co przejawia się w rozszerzonej dla większych tłumień charakterystyce selektywności.

Powyższe przykłady potwierdzają słuszność przeprowadzonej analizy i stosowanych metod obliczeniowych, wskazując jednocześnie na dużą zawilgość obliczeń przy wyznaczaniu charakterystyki selektywności metodą widmową. Związany z tą metodą pewien stopień przybliżenia uniemożliwia dokładne wyznaczenie przebiegu charakterystyki selektywności przy większych tłumieniach.

4. CHARAKTERYSTYKI SELEKTYWNOŚCI ODBIORNIKÓW SUPERREAKCYJNYCH PRACUJĄCYCH NIELINIOWO

Przeprowadzona analiza odnosiła się do układów superreakcyjnych, pracujących liniowo. Poniżej zostanie rozpatrzona możliwość zastosowania pewnych wyników tej analizy do układów pracujących nieliniowo.

Układy superrakcyjne pracujące nieliniowo służą z reguły do odbioru ciągłych sygnałów modulowanych i mają wewnętrzny układ detekcyjny, dający bezpośrednio na wyjściu sygnał małej częstotliwości. Ze względu na prawie stałą wartość maksymalnego poziomu impulsów drgań wielkiej

częstotliwości w układzie nieliniowym, nie można określić charakterystyki selektywności układu w taki sposób jak dla pracy liniowej.

Jak wiadomo, w nieobecności sygnału początkowy poziom drgań jest określony wartością napięcia szumu na wejściu układu. Na wyjściu otrzymuje się szum m. cz. charakterystyczny dla odbiorników superreakcyjnych pracujących nieliniowo. Z chwilą pojawienia się sygnału poziom szumu na wyjściu zmniejsza się w zależności od stosunku sygnału do szumu na wejściu odbiornika. Definiując charakterystykę selektywności układu odbiorczego jako zależność poziomu sygnału wejściowego od częstotliwości, który daje stały efekt wyjściowy, otrzymuje się prosty sposób wyznaczania charakterystyki selektywności układu nieliniowego. Przykładając na wejściu niemodulowany sygnał w. cz. o takiej wartości, aby zmniejszyć poziom szumu m. cz. na wyjściu o pewien stosunek, na przykład dziesięciokrotnie, określa się wartość tego sygnału w zależności od częstotliwości. Otrzymana zależność przedstawia charakterystykę selektywności układu odbiorczego.

Z analizy dla pracy liniowej wiadomo, że całkowite wzmocnienie układu dzieli się na wzmocnienie reakcyjne, zależne od częstotliwości sygnału, i wzmocnienie superreakcyjne, niezależne od częstotliwości sygnału. Ponieważ wzmocnienie superreakcyjne, nie zależy od częstotliwości sygnału i jednakowo działa na szum i sygnał, dla otrzymania stałego efektu wyjściowego m. cz. należy utrzymać stały stosunek sygnału do szumu po wzmocnieniu reakcyjnym.

Układ nieliniowy w początkowej części okresu wygaszania pracuje jako układ liniowy. Praca staje się nieliniowa dopiero przy stosunkowo dużym poziomie drgań, gdy sygnał zewnętrzny nie ma praktycznie żadnego wpływu na proces zachodzący w układzie superreakcyjnym. Stąd wynika, że wzmocnienie reakcyjne i selektywność układu nieliniowego są określone w liniowym okresie pracy układu i dla ich wyznaczenia można stosować te same wzory co dla rodzaju liniowego. A więc wszystkie rozważania dotyczące odkształceń charakterystyki selektywności przeprowadzone dla pracy liniowej, odnoszą się również do pracy nieliniowej.

* * *

Praca niniejsza stanowi zasadniczą część dysertacji kandydackiej złożonej na Wydziale Łączności Politechniki Warszawskiej.

Autor pragnie w tym miejscu złożyć gorące podziękowanie Prof. Drowi Inż. J. Groszkowskiemu — Kierownikowi Naukowemu Aspirantury za opiekę w czasie prowadzenia pracy i za krytyczne przejrzenie rękopisu. Dyrekcji Przemysłowego Instytutu Telekomunikacyjnego autor jest zobowiązany za zezwolenie na wykonanie części doświadczalnej pracy na terenie Instytutu, a Mgrowi Inż. J. Kroszczyńskiemu dziękuje za udział w dyskusji na temat analizy widmowej.

WYKAZ LITERATURY

1. Bloch A.: *Modulation theory*. J. I. E. E. Pt. III, vol. 91, s. 31.
2. Bradley W. E.: *Superregenerative detection theory*. Electronics vol. 21, nr. 9, s. 96.
3. Charkiewicz A. A.: *Spektry i analiz*. Moskwa 1952.
4. Fichtiengolc G. M.: *Kurs differencjalnowo i intiegralnowo isczislenija*. Tom III. Moskwa 1952, s. 642.
5. Groszkowski J.: *Generacja i stabilizacja częstotliwości*. Warszawa 1947, s. 79.
6. Hazeltine A., Richman D., Loughlin B. D.: *Superregenerator design*. Electronics, vol. 21, nr 9, s. 99.
7. Pogorzelski W.: *Analiza matematyczna*. T. III, Warszawa 1949, s. 159.
8. Richman D.: *Optimum selectivity in superregenerators*. Proceeding of the National Electronics Conference, November 4—6, 1948.
9. Riebman L.: *Superregenerative amplifier theory*. PIRE, vol. 37, nr 1, s. 29.
10. Van der Pol B.: *The fundamental principles of frequency modulation*. J. I. E. E. Pt. III, vol. 93, s. 153.
11. Whitehead P. R.: *Superregenerative receivers*. Cambridge 1950.
12. Wiśniewski F.: *Obliczanie układów superreakcyjnych*. Prace P. I. T. nr 16,

Dodatek

DOŚWIADCZALNE STWIERDZENIE WPŁYWU POSZCZEGÓLNYCH CZYNNIKÓW
NA WIELKOŚĆ ODKSZTAŁCANIA CHARAKTERYSTYKI SELEKTYWNOŚCI

1. METODY POMIAROWE

1.1. Układ pomiarowy

Celem doświadczeń było uchwycenie poszczególnych czynników wpływających na niesymetrię, jak również określenie ogólnych prawideł, którym podlega samo zjawisko. Doświadczenia przeprowadzono dla stałej częstotliwości sygnału a zmiennej częstotliwości układu odbiorczego. Pomiaru wykonywano na dwóch modelach pracujących w układzie Hartleya i Colpittsa na częstotliwości około 500 kHz.

W układzie pomiarowym jak na rys. 2 przeprowadzono szereg pomiarów, przy których za zmienny parametr kolejno uważano:

- a) wartość i znak zmiany częstotliwości układu odbiorczego,
- b) wartość części rzeczywistej współczynnika sprzężenia zwrotnego,
- c) kształt napięcia wygaszającego,
- d) rodzaj układu generacyjnego.

1.2. Pomiar charakterystyki selektywności

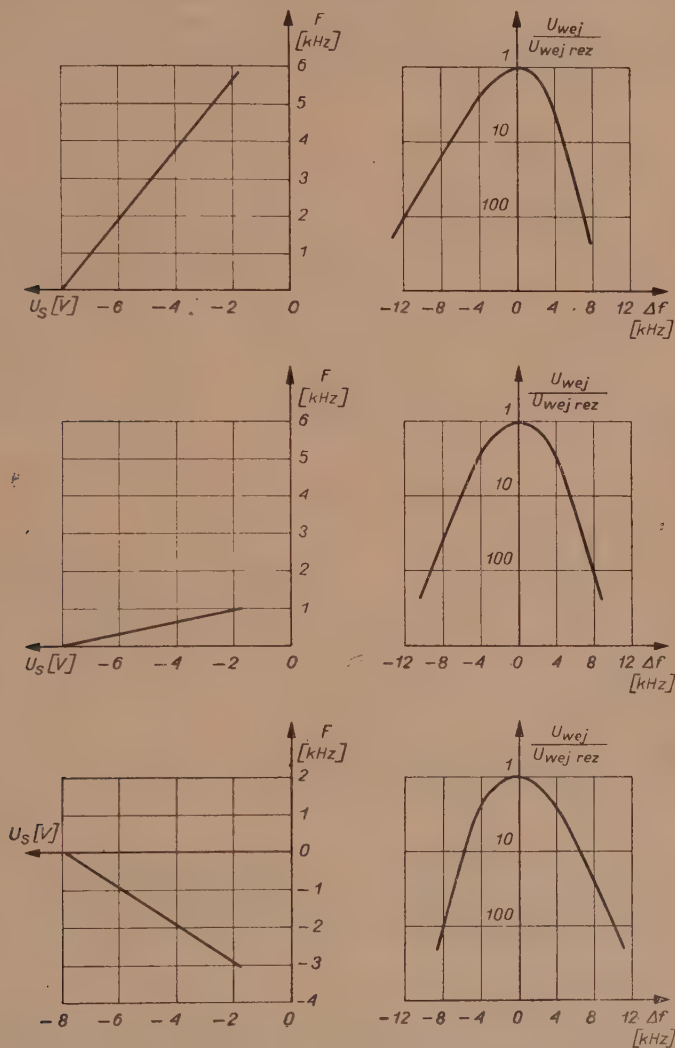
Pomiar charakterystyki selektywności przeprowadzono w identyczny sposób jak opisany w rozdziale 3.2.2.

1.3. Pomiar charakterystyki zmian częstotliwości

Do badanego odbiornika dołączono tak dobrany opór rzeczywisty, aby uzyskać przesunięcie punktu powstawania drgań układu w okolice zera napięcia siatkowego. Opór dołączono między anodę i katodę lampy, ponieważ takie dołączenie nie powoduje zmiany wartości współczynnika sprzężenia zwrotnego. Następnie, za pomocą możliwie małego sygnału wejściowego, określono częstotliwość rezonansową układu odbiorczego w funkcji napięcia siatkowego.

1.4. Pomiar charakterystyk selektywności dla różnych zmian częstotliwości układu odbiorczego

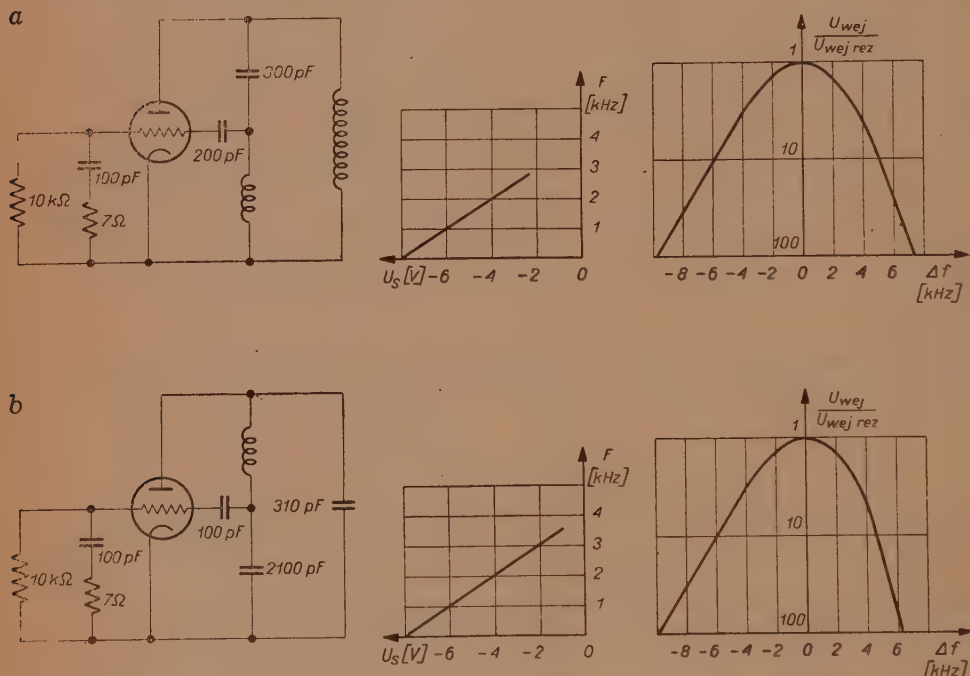
Dla wywołania zmian częstotliwości układu odbiorczego w funkcji napięcia siatkowego wykorzystano zjawisko reaktancyjnego działania lampy, które występuje w przypadku zespolonej wartości współczynnika sprzężenia zwrotnego (patrz 3. 1.).



Rys. 9. Zależność niesymetrii od charakterystyki zmian częstotliwości.

Przez zmianę wartości i znaku urojonej części współczynnika sprzężenia zwrotnego uzyskiwano różne co do wartości i kierunku zmiany częstotliwości układu odbiorczego w funkcji napięcia siatkowego. Zmianę wartości części urojonej współczynnika sprzężenia zwrotnego otrzymywano przez zmniejszenie lub zwiększenie

wartości oporu bocznikującego pojemność siatka-katoda. Ponieważ część rzeczywista współczynnika sprzężenia zwrotnego była znacznie większa od części urojonej, można nie uwzględniać wpływu zmiany oporu bocznikującego pojemność C_{kt} na wartość części rzeczywistej. Zmianę znaku części urojonej uzyskiwano w tym samym układzie Hartleya dołączając opory raz równolegle do pojemności C_{sk} , drugi raz szeregowo z indukcyjnością sprzęgającą i tak każdorazowo dobierając wartość obciążenia, aby uzyskać stały punkt wzbudzenia się drgań. Pomiary charakterystyk selektywności dla różnych charakterystyk zmian częstotliwości układu odbiorczego były wykonywane przy zadanym kształcie, wartości i częstotliwości napięcia wygaszającego.



Rys. 10. Niezależność niesymetrii od rodzaju układu generacyjnego.

a. układ Hartleya, b. układ Colpittsa.

1.5. Pomiar charakterystyk selektywności dla różnych kształtów napięcia wygaszającego

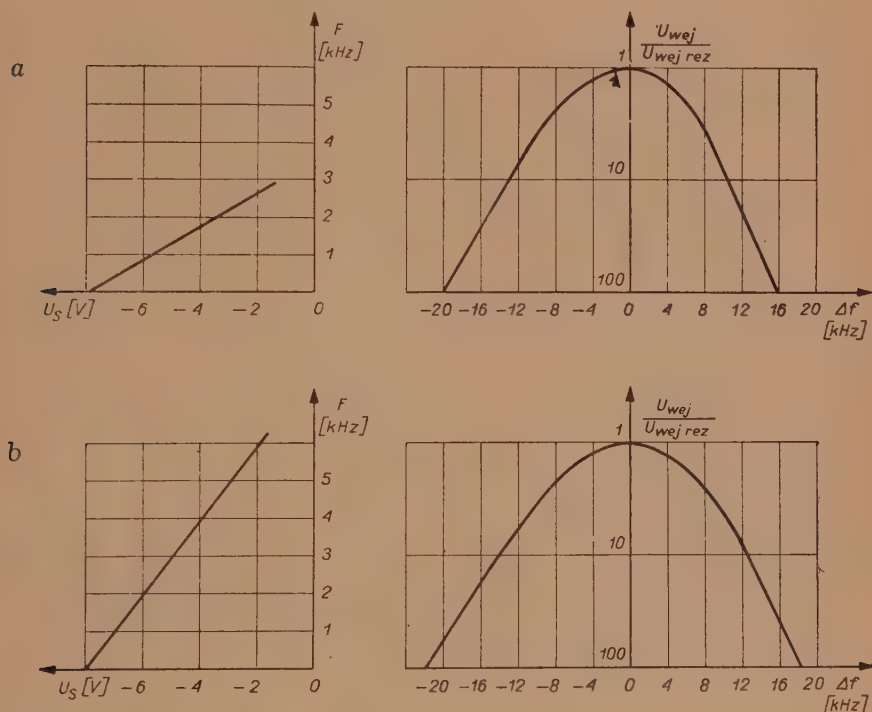
Do układu odbiorczego o stałych parametrach przykładano napięcie wygaszające o różnych kształtach i wartościach, lecz o tej samej częstotliwości i tak każdorazowo dobierano wartość stałego ujemnego napięcia siatkowego, aby otrzymać tę samą wartość wzmocnienia układu.

2. Dyskusja wyników doświadczalnych

Z przeprowadzonych doświadczeń wynika, że niesymetria krzywej selektywności nie zależy od rodzaju układu generacyjnego, lecz zależy od charakterystyki zmian częstotliwości układu odbiorczego, wartości części rzeczywistej współczynnika sprzężenia zwrotnego oraz od kształtu napięcia wygaszającego.

Dla zadanego napięcia wygaszającego i stałej wartości części rzeczywistej współczynnika sprzężenia zwrotnego niesymetria jest tym większa, im większa jest zmiana częstotliwości układu w okresie narastania drgań. Odwrócenie kierunku zmian częstotliwości powoduje zmianę niesymetrii, to znaczy, jeżeli charakterystyka była bardziej rozciągnięta w kierunku mniejszych częstotliwości, to po zmianie jest bardziej rozciągnięta w kierunku częstotliwości większych.

Wzrost części rzeczywistej współczynnika sprzężenia zwrotnego wymaga odpowiedniego wzrostu wartości zmian częstotliwości układu odbiorczego, dla otrzymania tej samej procentowej wartości niesymetrii.



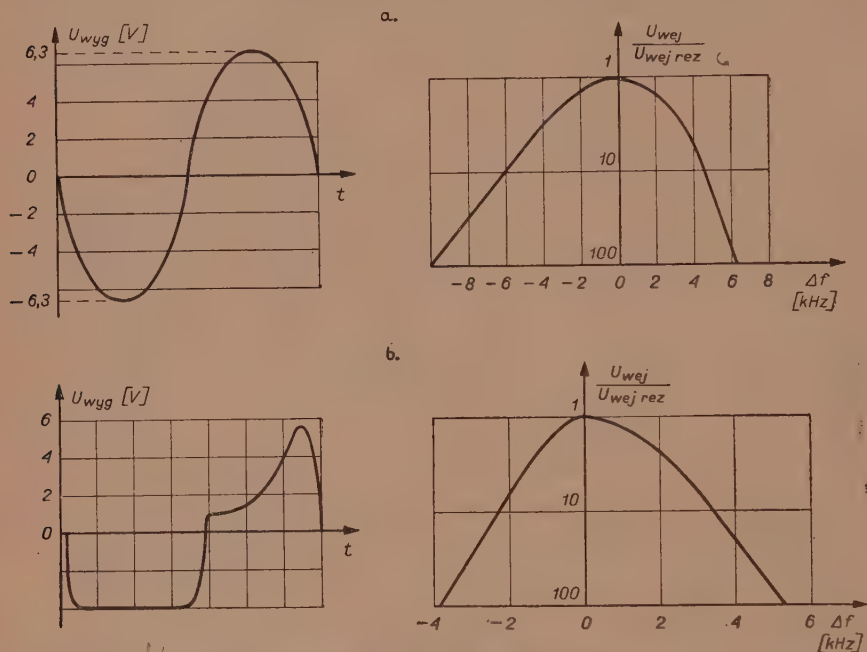
Rys. 11. Zależność niesymetrii od charakterystyki zmian częstotliwości i współczynnika sprzężenia zwrotnego. a. układ nieobciążony — słabo sprzęgnięty, b. układ obciążony — silnie sprzęgnięty.

Szczególny wpływ na niesymetrię ma kształt napięcia wygaszającego. Pozostawienie innych parametrów układów bez zmian, a zmienianie tylko samego kształtu napięcia wygaszającego wykazuje, że w zależności od tego czy napięcie wygaszające ma w czasie narastania drgań kształt wklęsły czy wypukły, krzywa selektywności jest bardziej rozszerzona z jednej lub drugiej strony swego wierzchołka.

Przedstawiona w poprzednich rozdziałach teoria jest w całkowitej zgodności z powyższymi wynikami doświadczalnymi i tłumaczy je w następujący sposób.

Pierwsza grupa doświadczeń obejmowała zależność niesymetrii charakterystyki selektywności od wartości zmian częstotliwości układu odbiorczego. Doświadczenia

wykonano dla układu o zadanej wartości części rzeczywistej współczynnika sprzężenia zwrotnego i o zadanym napięciu wygaszającym. Oznacza to, że moduł funkcji czułości był wartością stałą. Zwiększenie zmian częstotliwości jest równoznaczne powiększeniu części urojonej funkcji czułości w stosunku do części rzeczywistej i powoduje wzrost wartości niesymetrii charakterystyki. Odwrócenie kierunku zmian częstotliwości zmienia znak niesymetrii, ponieważ równocześnie odwraca niesymetrię



Rys. 12. Zależność niesymetrii od kształtu napięcia wygaszającego.
a. napięcie wygaszające o kształcie wypukłym, b. napięcie wygaszające o kształcie wklęsłym.

składowej rzeczywistej i składowej urojonej transformaty. Przy stałej częstotliwości układu funkcja czułości ma wartość rzeczywistą i daje symetryczną charakterystykę selektywności.

Dalsze doświadczenia wykazały, że dla utrzymania stałej procentowej wartości niesymetrii należy równocześnie ze wzrostem części rzeczywistej współczynnika sprzężenia zwrotnego zwiększyć wartość zmiany częstotliwości układu. Na podstawie przedstawionej teorii zależność ta jest oczywista. Zwiększenie części rzeczywistej współczynnika sprzężenia zwrotnego jest równoznaczne zmniejszeniu szerokości przebiegu określonego funkcją czułości. Dla otrzymania tej samej procentowej niesymetrii, konieczne jest zachowanie stałego stosunku między częścią rzeczywistą a urojoną funkcji czułości, co wymaga odpowiedniego wzrostu wartości zmian częstotliwości.

Zmiana układu generacyjnego nie ma wpływu na niesymetrię, gdyż nie zmienia kształtu funkcji czułości.

Kształt napięcia wygaszającego określa przebieg funkcji przewodności i przebieg zmian częstotliwości układu. Zmiana kształtu napięcia wygaszającego z wklęsłego

na wypukły, przy liniowych w przybliżeniu statycznych charakterystykach zmian przewodności $g(u_s)$ i częstotliwości $f(u_s)$, powoduje odwrócenie znaku niesymetrii modułu funkcji czułości, jak również niesymetrii składowej rzeczywistej i urojonej transformaty, co daje zmianę znaku niesymetrii charakterystyki selektywności.

Ф. ВИСЬНЕВСКИ

АНАЛИЗ СУПЕРРЕГЕНЕРАТИВНОЙ СИСТЕМЫ С ЗАТУХАНИЕМ ВЫЗЫВАЮЩИМ МОДУЛЯЦИЮ ЧАСТОТЫ

Резюме

При измерениях характеристик избирательности суперреактивных систем наблюдается явление частого отклонения этих характеристик от теоретических кривых, рассчитанных на основании известных в литературе формул, причём их форма становится явно несимметричной.

Явление это было замечено несколькими авторами, сообщающими, что оно выступает в случае изменения частоты приёмной системы в пределах её чувствительности. Для подтверждения и дополнения этого наблюдения был проведён ряд опытов, результаты которых могут быть изложены следующим образом:

Необходимым условием выступления асимметрии характеристики избирательности является изменение частоты системы в периоде её чувствительности. Асимметрия характеристики системы зависит одновременно:

- а) от характеристики изменений частоты приёмной системы,
- б) от значения действительной части коэффициента обратной связи и
- в) от формы суперрегенеративного напряжения.

В труде представлен анализ суперрегенеративной системы с затуханием, вызывающим модуляцию частоты. Рассматриваемая была система (рис. 1), заключающая источник сигнала в виде генератора тока, соединённого с параллельным резонансным контуром, которого проводимость и ёмкость являются функциями времени. Предполагая линейную работу системы — было составлено дифференциальное уравнение следующей формы:

$$u'c + uc' + gu + \frac{1}{L} \int u dt = i$$

которое было решено принимая, что

$$\left| \left(\frac{g}{2c} \right)' - \left(\frac{g}{2c} \right)^2 \right| \ll \omega_0^2 \quad \text{для всякого} \quad \frac{g}{2c}$$

и

$$\omega' \ll \omega_0^2$$

Применяя затем ряд преобразований и упрощений, приемлемых для типичных условий работы суперрегенеративных систем, получено решение

$$u(t_2) = \left| \frac{I}{2C_0} \frac{\omega_s}{\omega_0} \exp \left(\frac{1}{2} \int_0^{t_1} p_0 dt \right) \int_{t_1}^{t_2} \exp \left(\frac{1}{2} \int_0^{\tau} p_0 dt \right) \exp [-j\psi(\tau)] \exp (-j\Omega\tau) d\tau \right|$$

где

$$p_0 = \frac{g}{C_0}; \quad \Omega = \omega_0 - \omega_s, \quad \psi(\tau) = \int_0^{\tau} \omega d\tau - \omega_0 \tau$$

Приведенное решение имеет более общий характер, чем известные до сих пор другие решения, так как оно относится к системам как с переменной, так и с постоянной частотой. Решение это позволяет определить *амплитуду* напряжения на контуре, усиление системы и характеристику избирательности. Из проведенного анализа следует, что изменение резонансной частоты приемного контура в периоде его чувствительности имеет свой математический эквивалент в комплексном значении импульса чувствительности. Характеристика избирательности суперрегенеративной системы связана с импульсом чувствительности таким же образом, как спектр с функцией, которой он отвечает, поэтому — для более точного определения влияния изменения частоты на характеристику избирательности системы — проведено было обсуждение свойств спектра, функция которого имеет комплексное значение.

Применяя распределение как функции так и спектра на части чётные и нечётные, а также на части действительные и мнимые, представлен ряд теорем исчерпывающих теоретически возможные случаи.

Основываясь на этих теоремах, можно сделать следующие выводы:

Если изменения проводимости и частоты происходят симметрично по отношению к моменту возбуждения собственных колебаний в системе, то действительная и мнимая составляющие импульса чувствительности являются чётными функциями. Характеристика избирательности такой системы будет симметрична, но возможна её деформация и расширение в сравнении с характеристикой, соответствующей системе с постоянной частотой. Если изменения проводимости и частоты не симметричны по отношению к моменту возбуждения собственных колебаний системы, то действительная или мнимая составляющая импульса чувствительности, или же обе эти составляющие имеют одновременно чётные и нечётные части. Характеристика избирательности такой системы будет несимметрична за исключением особых случаев, определённых в предположениях теоремы (4). Возможны также дополнительная деформация и расширение по отношению к характеристике системы с постоянной частотой.

Для двух экспериментальных систем были сравнены результаты расчётов с результатами измерений. Полученное согласие этих результатов подтверждает правильность проведенного анализа, а также применяемых методов расчёта.

В конечной части труда доказана возможность применения формул полученных из анализа линейной системы к определению характеристики избирательности нелинейной системы.

В дополнении представлены опытные результаты влияния отдельных факторов на величину деформации характеристики системы и доказано согласие полученных результатов с приведенной теорией.

F. WIŚNIEWSKI

THE ANALYSIS OF THE SUPERREGENERATIVE CIRCUIT WITH QUENCHING CAUSING FREQUENCY MODULATION

Summary

When measuring the selectivity characteristics of the supergenerative circuits one may observe, that in many cases these characteristics differ from the theoretical curves calculated by equations known in the literature, and get a clearly asymmetrical shape.

This phenomenon was announced by several authors, who also noted that it appears in the case of the change of the receiver circuit frequency during its sensitivity period. In order to confirm and to complete these observations a series of experiments were performed, the results of which may be recapitulated in the following way: The necessary condition for the appearance of selectivity characteristic asymmetry is the change of the circuit frequency in its sensitivity period. The circuit asymmetry depends simultaneously on:

- a. the characteristics of the frequency changes of the receiver circuit,
 - b. the value of the real part of the feedback factor,
- and
- c. the shape of the quench voltage.

In this paper an analysis of the superregenerative circuit with quenching causing frequency modulation is shown. A circuit (Fig. 1) was considered, which contains the signal source in the form of a current generator connected with a parallel resonance circuit, the conductance and capacity of which are a function of time. Assuming a linear law of circuit performance, the following differential equation has been derived:

$$u'c + uc' + gu + \frac{1}{L} \int u dt = i$$

which is solved assuming that

$$\left| \left(\frac{g}{2c} \right)' - \left(\frac{g}{2c} \right)^2 \right| \ll \omega_0^2 \quad \text{for every} \quad \frac{g}{2c}$$

and

$$\omega' \ll \omega_0^2$$

Applying then transformations and simplifications allowed for typical conditions of superregenerative circuits, a solution in the following form has been obtained

$$u(t_2) = \left| \frac{I}{2C_0} \frac{\omega_s}{\omega_0} \exp \left(\frac{1}{2} \int_0^{t_2} p_0 dt \right) \int_{t_1}^{t_2} \exp \left(\frac{1}{2} \int_0^{\tau} p_0 dt \right) \exp [-j\psi(\tau)] \exp (-j\Omega\tau) d\tau \right|$$

where

$$p_0 = \frac{g}{C_0}, \quad \Omega = \omega_0 - \omega_s, \quad \psi(\tau) = \int_0^{\tau} \omega d\tau - \omega_0 \tau$$

The above solution is of a more general character than any others known, because it refers to circuits of variable and constant frequency. This solution allows to determine the amplitude of the circuit voltage, as well as the voltage amplifications of the circuit and its selectivity characteristics.

The analysis shows, that the change of the receiver circuit resonance frequency in its sensitivity time interval has its mathematical correspondance in the complex value of the sensitivity function. Since the characteristic of the superregenerative selectivity circuit is as connected with the sensitivity function as Fourier's transformation with that function, to which it corresponds, a discussion has been carried out about the properties of the transformation, the function of which has a complex value. It was to determine the influence of the frequency change on the characteristic of the selectivity circuit. Having divided the functions and the transformations into even and odd parts, and into real and imaginary ones, a series of theorems were shown, that discuss all cases theoretically possible.

On the base of the above theorems one may conclude as follows: If the changes of the conductance and frequency are symmetrical respectively to the moment of zero damping, then the real and imaginary components of the sensitivity function are even functions. The selectivity characteristic of such circuit will be symmetrical, but its widening and deformation in comparison to the characteristic corresponding to a circuit of a constant frequency is possible. If the admittance and frequency changes are not symmetrical respectively to the moment of zero damping, then the real or imaginary components of the sensitivity function or both have even and odd parts simultaneously. The selectivity characteristic of such circuit will be asymmetrical (except the particular cases determined in the assumption of theorem (4). Also an additional deformation and widening is possible respectively to the circuit characteristic of a constant frequency.

The calculated results for two experimental circuits were compared with the measured results. The received agreement of the results proves that the performed analysis and applied calculation method are correct.

In the last part of the paper it is shown that the equations obtained from the linear mode analysis are applicable for determination of the selectivity characteristic of the logarithmic mode.

In the Appendix the experimental results of the influence of the separate factors upon the amount of the circuit characteristic deformation are shown, and the agreement of the obtained results and the given theory is proved.

538.56:535.3

K. BOCHENEK, J. PLEBAŃSKI

O metodach optyki mikrofalowej

Rękopis dostarczono 18. 11. 1955

Zostały zaproponowane i omówione dwie metody aproksymacji pól elektromagnetycznych, będące rozwinięciem metod optyki geometrycznej. Metody te mają na celu ułatwienie analizy zagadnień związanych z techniką mikrofalową i zostały zilustrowane przykładami obliczeń.

1. WSTĘP

Technika anten mikrofalowych posługuje się w ostatnich latach coraz szerzej metodami optycznymi. Rozwojowi zastosowań tych metod nie można się dziwić, jest on spowodowany ich prostotą, poglądowością oraz stosunkowo dużą efektywnością. Z drugiej jednak strony specyficzny rząd wielkości parametrów urządzeń mikrofalowych powoduje, iż klasyczne prawa i metody optyki geometrycznej, bez zarzutu spełniające swe zadania w stosunku do reflektorów, soczewek i podobnych układów dla fal świetlnych ($\lambda \approx 4000 \text{ \AA}$), w przypadku układów mikrofalowych — o mniejszych z reguły wymiarach w stosunku do długości fali ($\lambda \approx 10 \text{ cm}$) — okazują się mniej ściśle. Stosowalność tych metod podlega więc znacznie większym ograniczeniom.

Z tego względu poświęcono ostatnio dużo uwagi poszukiwaniu kompromisu między optyką geometryczną a falową; kompromis taki polegałby na opracowaniu metod zachowujących częściowo prostotę metod optycznych i pozwalających zarazem na uchwycenie najistotniejszych odchyleń od praw klasycznej optyki geometrycznej. Badania w tej dziedzinie prowadzone są tak szeroko, iż coraz częściej na łamach prasy fachowej pojawia się termin optyka mikrofalowa jako określenie nowej, specyficznej dyscypliny.

Przedstawiony obraz mógłby jednak wywołać wrażenie, że metody optyczne ze względu na swe zalety stanowią narzędzie, które ma zastąpić inne metody analizy pola, a w szczególności metody teorii dyfrakcji (wzory Kirchhoffa, Kottlera itp.). Wniosek taki byłby zasadniczo fałszywy i nie leżał w intencjach autorów.

W rzeczywistości — biorąc z grubsza — metody optyczne stanowią nieźle narzędzie analizy pola w bezpośrednim otoczeniu apertury anteny, natomiast w stosunku do pola w strefie dalekiej stanowią jedynie narzędzie pomocnicze, niekiedy pożyteczne przez swą prostotę.

Celem pracy jest rzucić nieco światła na stosunek praw klasycznej optyki do ścisłej teorii opartej na równaniu falowym. W szczególności chodzi o możliwość stosowania optycznej interpretacji niektórych zależności uwzględniających ściślej niż optyka geometryczna zjawiska opisywane przez równanie falowe. Będziemy starali się rozwinąć przy tym dwie alternatywne metody pozwalające podejść efektywnie do ilościowego opisu zjawisk mikrofalowych, ściślejszego niż w klasycznych metodach optycznych.

Praca nie pretenduje do zupełności i jest jedynie szkicem pewnego, o ile wiemy, nowego podejścia do zagadnienia, które zresztą wydaje się intuicyjnie bardzo naturalne.

2. POSTAWIENIE ZAGADNIENIA

W pracy tej będziemy się zajmować wyłącznie polem skalarnym $U=U(\mathbf{x},t)$ rządzonego przez równanie falowe

$$\square U \equiv \left(\Delta - \frac{1}{v^2} \frac{\partial^2}{\partial t^2} \right) U = 0, \quad (1)$$

gdzie $v = \frac{1}{\sqrt{\epsilon\mu}} = v(\mathbf{x})$ jest prędkością propagacji fal (czoła fali, prędkością jednoznaczności), która w ogólnym przypadku ośrodka niejednorodnego jest daną określoną funkcją punktu przestrzeni $\mathbf{x}$. Ujęcie takie, jak wiadomo, jest właściwe w przypadku, gdy postawiony problem brzegowy dotyczy jednej ze współrzędnych prostokątnych natężenia pola.

Ujęcie wektorowe jest potrzebne wtedy, gdy chcemy uwzględnić związki pomiędzy poszczególnymi współrzędnymi, na przykład przy wyprowadzeniu praw odbicia i załamania. W takich przypadkach ujęcie skalarne jest zbyt ogólne dla osiągnięcia zamierzonych rezultatów. Wektorowa postać omawianych w tej pracy metod pozwoliłaby również na uzyskanie dalszych informacji o badanych zjawiskach. Sprawie tej zamierzamy poświęcić dalszą publikację.

W przypadku najbardziej w praktyce interesujących stanów ustalonych U dane jest jako $U = \Re_c [\varphi(\mathbf{x}) \exp i\omega t]$, gdzie zespolona funkcja $\varphi(\mathbf{x})$ spełnia równanie

$$(\Delta + k^2) \varphi = 0,$$

w którym

$$k = \frac{\omega}{v} = k(\mathbf{x}) = k_0 n(\mathbf{x}), \quad (2)$$

gdzie

$$k_0 = \frac{\omega}{c},$$

$n = \frac{c}{v}$ jest współczynnikiem załamania w ośrodku niejednorodnym.

Zagadnienia praktyki wiążą z równaniem (2) następujący problem matematyczny: znaleźć $\varphi(\mathbf{x})$ spełniające równanie (2) oraz warunek brzegowy $\varphi(\mathbf{x}) = \bar{\varphi}(\mathbf{x}) = \bar{A}(\mathbf{x}) \exp i\bar{\vartheta}(\mathbf{x})$ na pewnej powierzchni Σ . Innymi słowy, przy danej amplitudzie i fazie φ spełniającego (2) na Σ , należy znaleźć amplitudę i fazę rozwiązania w całym interesującym nas obszarze przestrzennym.

Na czym polega przejście do przybliżenia optyki geometrycznej w naszym zagadnieniu? Podstawmy do równania (2) $\varphi = A \exp \{ik_0 L\}$ i rozdzielmy równanie na części: rzeczywistą i urojoną. W rezultacie otrzymamy¹

$$(\nabla L)^2 = n^2 + \frac{1}{k_0^2} \frac{\Delta A}{A}, \quad (3a)$$

$$\nabla(A^2 \nabla L) = 0. \quad (3b)$$

Problem całkowania równań (3a) i (3b) przy warunkach brzegowych $A = \bar{A}$, $L = \bar{\vartheta}/k_0$ na Σ jest oczywiście równoważny interesującemu nas problemowi brzegowemu dla równania Helmholtza (2).

Przybliżenie typu optyki geometrycznej w tym problemie otrzymuje się gdy $n^2 \gg \frac{1}{k_0^2} \frac{\Delta A}{A}$, przez zastąpienie układu (3a), (3b) układem

$$(\nabla L)^2 = n^2, \quad (4a)$$

$$\nabla(A^2 \nabla L) = 0. \quad (4b)$$

Jeśli scałkujemy ten układ przy warunkach brzegowych $A = \bar{A}$, $L = \frac{\bar{\vartheta}}{k_0}$ na Σ , to zbudowane z tak otrzymanych L i A

$$U = \mathcal{R}_e A \exp \{i(\omega t + k_0 L)\}$$

będzie przybliżonym rozwiązaniem odpowiadającym przybliżeniu optyki geometrycznej.

¹ Jeżeliby w (1) podstawić $A e^{iS}$ i zażądać, aby przy rzeczywistych $A(\mathbf{x}, t)$ i $S(\mathbf{x}, t)$ równanie (1) było spełnione, to w wyniku otrzyma się równania

$$\frac{\square A}{A} = (\nabla S)^2 - \frac{1}{v^2} \left(\frac{\partial S}{\partial t} \right)^2, \quad \nabla(A^2 \nabla S) + \frac{\partial}{\partial t} \left(-\frac{1}{v^2} \frac{\partial S}{\partial t} A^2 \right) = 0.$$

Podstawienie to i otrzymane za jego pomocą równania wymagają oczywiście odrębnej interpretacji i dyskusji. Specjalizując $A = A(\mathbf{x})$ i $S = \omega t + k_0 L(\mathbf{x})$ otrzymamy równania (3) z tekstu.

Jaki sens fizyczny mają równania (4a) (równanie eikonalne) i (4b). Sens fizyczny równania (4a) jest bardzo prosty. Rozważmy falę

$$U = \mathcal{R}_e A \exp \{i(\omega t + k_0 L)\}$$

w otoczeniu ustalonego punktu $\mathbf{x}_0$. W założeniu wolnej zmienności $A(\mathbf{x})$ z dokładnością do małych rzędów wyższych U możemy napisać w otoczeniu tego punktu jako

$$U = \text{Re } A(\mathbf{x}) \exp \{i[\omega t + k_0 \nabla L(\mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0) + k_0 L(\mathbf{x}_0)]\}.$$

Porównując to wyrażenie z wyrażeniem dla fali płaskiej

$$U = \text{Re } A \exp \{i[\omega t - \mathbf{k}(\mathbf{x} - \mathbf{x}_0) + \delta]\}$$

widzimy, że lokalnie nasza fala zachowuje się jak fala płaska o fazie $\delta = k_0 L(\mathbf{x}_0)$, amplitudzie $A = A(\mathbf{x}_0)$ i wektorze falowym $\mathbf{k} = -k_0 \nabla L(\mathbf{x})$; lokalną prędkością fazową tej fali będzie zatem

$$v_f = \frac{\omega}{k_0 |\nabla L(\mathbf{x}_0)|} = \frac{c}{|\nabla L(\mathbf{x}_0)|},$$

stosunek prędkości światła w próżni c do v_f wynosi więc $|\nabla L| = \frac{c}{v_f}$,

Równanie (4a) wyraża fizycznie fakt, iż niezależnie od zachowania się amplitudy w całej przestrzeni, stosunek prędkości światła w próżni do

lokalnej prędkości fazowej wynosi $n = n(\mathbf{x}) = \sqrt{\frac{\epsilon'(\mathbf{x})}{\epsilon_0 \mu_0}}$, jest zatem równy danej z góry funkcji ustalonej dla ośrodka. Równanie (4b), identyczne zresztą z (3b), daje znane prawo propagacji amplitudy w optyce geometrycznej wyznaczone wyłącznie przez zmienność fazy. Optyka, obok pojęcia powierzchni $\text{é} \text{quifazowych}$ $L = \text{const}$, wprowadza pojęcie promieni jako linii ortogonalnych do rodziny powierzchni $L = \text{const}$ (to jest całek układu równań zwyczajnych $\nabla L \times d\mathbf{x} = 0$). Równanie (4b) można interpretować jako wyrażające fakt, że A^2 rozchodzi się wzdłuż promieni.

Charakterystyczną cechą tego obrazu optyki geometrycznej jest ode-
rwanie zmienności fazy od zachowania się amplitudy. Oczywiście, wy-
łącznie równanie (4a) i warunek $L = \frac{\partial}{\partial k_0}$ na Σ są zależnościami, które mają
wyznaczać L w całym interesującym nas obszarze.

Sądzimy jednak — i to jest pierwsza zasadnicza idea naszej pracy — iż aby uzyskać prosty obraz optyki geometrycznej, nie ma konieczności uniezależniania fazy od amplitudy.

Wróćmy do ścisłych równań (3a) i (3b):

$$(\nabla L)^2 = n^2 + \frac{1}{k_0^2} \frac{\Delta A}{A}, \quad \nabla(A^2 \nabla L) = 0.$$

Lokalną prędkością fali $U = \text{Re } A \exp \{i(k_0 L - \omega t)\}$ będziemy nazywali $v_f = \frac{c}{|\nabla L|}$. Wielkość ta jest mierzalna i przy A wolnozmiennym posiada przejrzystą interpretację podaną wyżej. Pierwsze z naszych ścisłych równań orzeka więc fizycznie, że $N = c/v_f$ dane jest jako

$$N = \sqrt{n^2 + \frac{1}{k_0^2} \frac{\nabla A}{A}}. \quad (5)$$

Powyższe sugeruje następującą interpretację układu (3a), (3b) równoważnego równaniu Helmholtza:

a. A i L — spełniające (3a) i (3b) nazywamy odpowiednio amplitudą i eikonalem,

b. linie wektorowe pola ∇L nazywamy promieniami,

c. $N = \sqrt{n^2 + \frac{1}{k_0^2} \frac{\Delta A}{A}}$ — uważamy za współczynnik załamania, który zależy nie tylko od x wskutek niejednorodności ośrodka, lecz również za pośrednictwem zmienności amplitudy w przestrzeni.

Interpretacja ta wykazuje dużą analogię z interpretacją mechaniki kwantowej D. Bohma [1], gdzie, podobnie jak my, przez wprowadzenie współczynnika załamania N czynimy z teorii falowej swojego rodzaju optykę geometryczną; tak D. Bohm przez wprowadzenie potencjału kwantowego sprowadza mechanikę falową do swojego rodzaju deterministycznej mechaniki². O ile jednak Bohm budował swoją interpretację ze względów filozoficznych, o tyle nasza interpretacja ma na celu osiągnięcie konkretnych rezultatów rachunkowych. Nie zamierzamy bynajmniej rezygnować z pojęcia fali na korzyść wyżej arbitralnie określonych promieni. Co więcej, sądzimy, że interpretacja zawarta w punktach a. b. i c. traci sens i znaczenie praktyczne, jeżeli $\frac{1}{k_0^2} \left| \frac{\Delta A}{A} \right| \approx n^2$, przy czym amplituda jest funkcją szybkozmienną. Nam zaś przede wszystkim chodzi o pewne korzyści praktyczne, jakie może przynieść ta interpretacja. Sądzimy, iż gdy:

1. $\frac{1}{k_0^2} \left| \frac{\Delta A}{A} \right|$ jest małe w porównaniu z n^2 , ale nie do zaniedbania,

2. A jest stosunkowo wolnozmienną funkcją punktu.

Wówczas N jest stosunkowo wolnozmienną funkcją i ścisła faza $k_0 L$ zachowuje się podobnie spokojnie i regularnie jak eikonale w optyce geometrycznej w ośrodku niejednorodnym. Promienie te posiadają

² Uwagę, iż równanie falowe można by interpretować podobnie jak Bohm to czyni z równaniem Schrödingera, jako zarzut pod adresem interpretacji Bohma podał prof. Szczeniowski na konferencji fizyków w Spale (1954).

oczywiście odstępstwa od promieni w przypadku klasycznej optyki, lecz bieżną dostatecznie regularnie i bez fluktuacji, aby móc być pożytecznym narzędziem w analizie pola i ewentualnym projektowaniu własności pola w otoczeniu na przykład radiatorów. Zarazem uwzględniając w N^2 wyraz $\frac{1}{k_0^2} \frac{\Delta A}{A}$, porzucamy sztuczny rozdział fazy od amplitudy

indukowany przez optykę geometryczną, według którego na przykład przebieg powierzchni equifazowych nie zależał od warunku początkowego na amplitudę A na powierzchni Σ , gdzie zadajemy warunki brzegowe.

Taka ścisła optyka oparta na interpretacji a. b. c. oczywiście niewiele byłaby warta w praktyce, gdyby nie poprzeć jej nową metodą rachunku, gdyż w przeciwnym razie, można by co najwyżej od wyprowadzonych równań (3) wrócić do równania Helmholtza, to znaczy do punktu wyjścia.

Dlatego też w rozdziale trzecim rozwinięto metodę przybliżonego rozwiązywania równania (3a), (3b), opartą na idei rozwinięcia na szereg w $\frac{1}{k_0}$.

Rachunki nasze będą jednak dość formalne; oczywiście, z matematycznego punktu widzenia metoda, którą rozwinieśmy, wymagałaby poparcia dowodem zbieżności. Podana wyżej interpretacja będzie pozwalała na interpretację wyniku rachunku każdego kroku przybliżenia za pomocą prostych pojęć optycznych. Dalej będą podane przykłady zastosowania tej metody i interpretacji.

Omówimy teraz drugą węzłową ideę pracy — alternatywną do pierwszej.

Otóż dobrze wiadomo, iż w mechanice kwantowej daje dobre wyniki, nawiązując jako do zerowego kroku przybliżenia do mechaniki klasycznej, metoda przybliżonego rozwiązywania równania Schrödingera, zwana metodą W. K. B. (Wentzel, Kramers, Brillouin) [2]. Metoda ta, z grubsza biorąc, polega na poszukiwaniu zespolonego rozwiązania równania

$$\left(-\frac{\hbar^2}{2m} \Delta + V(\mathbf{x}) - E \right) \Psi(\mathbf{x}) = 0$$

w postaci $\psi(\mathbf{x}) = \exp \frac{iS(\mathbf{r})}{\hbar}$, gdzie zespolone S rozwijalne jest na szereg potęgowy w $\frac{\hbar}{i}$, to jest

$$S = \sum_{n=0}^{\infty} \left(\frac{\hbar}{i} \right)^n S_n.$$

Aby uzyskać rozwiązanie problemu brzegowego dla równania Helmholtza (2), można próbować postąpić w analogiczny sposób.

W równaniu Helmholtza $(\Delta + k_0^2 n^2) \varphi$ (6)

będziemy poszukiwali rozwiązania w postaci

$$\varphi = \exp ik_0 A, \quad (7)$$

gdzie zespolone A rozwija się na szereg potęgowy w $\frac{1}{k_0}$, to jest

$$A = \sum_{n=0}^{\infty} \left(\frac{1}{ik_0} \right)^n {}_n A. \quad (8)$$

Współczynniki przy $\left(\frac{1}{ik_0} \right)^n$ oznaczamy ${}_n A$.

Jeżeli dokonamy podstawienia (7) w (6) i zażądamy, po uwzględnieniu (8), zachodzenia równania w każdym rzędzie w $\frac{1}{ik_0}$ to otrzymamy następujący ciąg równań:

$$(\nabla_0 A)^2 = n^2, \quad (9a)$$

$$\sum_{p=0}^n \nabla_p A \nabla_{n-p} A + A_{n-1} A = 0, \quad n \geq 1, \quad (9b)$$

Łatwo zauważyć, że jeżeli powyższy ciąg równań pierwszego rzędu na zespolone funkcje ${}_n A$ rozwiążemy przy warunkach początkowych

$$\begin{aligned} {}_0 A &= \frac{\bar{\partial}}{k_0} + \frac{1}{ik_0} \ln A, \\ \mathbf{x} \in \Sigma \rightarrow {}_n A &\equiv 0 \quad n \geq 1, \end{aligned} \quad (10)$$

to wówczas $\varphi = \exp \left[ik_0 \sum_{n=0}^{\infty} \left(\frac{1}{ik_0} \right)^n {}_n A \right]$ będzie, oczywiście w założeniu, że taka procedura jest zbieżna, rozwiązaniem problemu brzegowego dla równania Helmholtza.

Pewne wątpliwości może wywołać fakt, iż stosując metodę W.K.B rozwijamy rozwiązanie w określony sposób na szereg w $\frac{1}{k_0}$, nie przywiązujemy zaś znaczenia do faktu, iż w warunku brzegowym na Σ występuje $\frac{1}{k_0}$.

Łatwo jest jednak zdać sobie sprawę z możliwości takiego postępowania. W przypadku stanów ustalonych k_0 jest ustaloną konkretną liczbą. Wprowadzenie jej w warunek brzegowy (10) można traktować jako określoną definicję ${}_0 \bar{A}$ na Σ przez

$$\frac{\bar{\partial}}{k_0} + \frac{1}{ik_0} \ln \bar{A},$$

$$nA[(\nabla_0 L)^2 - n^2] + \sum_{q=i}^n n_{-q} A \sum_{p=0}^q \nabla_{q-p} L \nabla_p L = \Delta_{n-2} A, \quad n \geq 2 \geq 3.$$

Równania pochodne z równania (3b) dla $n \geq 2$ będą następujące:

$$\begin{aligned} 2\nabla_0 A \cdot \nabla_0 L + {}_0 A \Delta_0 L &= 0, \\ 2(\nabla_1 A \cdot \nabla_0 L + \nabla_0 A \cdot \nabla_1 L) + {}_1 A \Delta_0 L + {}_0 A \Delta_1 L &= 0, \\ &\dots \dots \dots \\ 2\nabla_n A \cdot \nabla_0 L + {}_n A \Delta_0 L + \sum_{p=1}^n (2\nabla_{n-p} A \cdot \nabla_p L + {}_{n-p} A \Delta_p L) &= 0. \end{aligned} \quad (12b)$$

Rozważając strukturę tych równań spostrzegamy łatwo, że o ile ${}_0 A \neq 0$, wówczas redukują się one na następujący ciąg par równań na pary niewiadomych funkcji ${}_n A$ i ${}_n L$

$$(\nabla_0 L)^2 - n^2 = 0, \quad \nabla({}_0 A^2 \nabla_0 L) = 0, \quad (13a)$$

$$\nabla_1 L \cdot \nabla_0 L = 0, \quad 2\nabla_1 A \cdot \nabla_0 L + {}_1 A \Delta_0 L + 3\nabla_0 A \cdot \nabla_1 L + {}_0 A \Delta_1 L = 0, \quad (13b)$$

dla $n \geq 2$

$$\begin{aligned} 2{}_0 A \nabla_n L \cdot \nabla_0 L = \Delta_{n-2} A - {}_0 A \sum_{p=1}^{n-1} \nabla_{n-1-p} L \cdot \nabla_p L + \\ - \sum_{q=1}^{n-1} {}_{n-1-q} A \sum_{p=0}^q \nabla_{q-p} L \cdot \nabla_p L, \quad (13c) \\ 2\nabla_n A \cdot \nabla_0 L + {}_n A \Delta_0 L = - \sum_{p=1}^n (2\nabla_{n-p} A \nabla_p L + {}_{n-p} A \Delta_p L). \end{aligned}$$

Charakterystyczna struktura równań pozwala rozwiązywać je sukcesywnie. Pierwsze z nich — to równanie pierwszego rzędu na ${}_0 L$. Skoro znajdziemy ${}_0 L$, to wyznaczenie ${}_0 A$ sprowadza się do całkowania drugiego równania pary (13a), które jest na ${}_0 A$ równaniem liniowym pierwszego rzędu. Następnie z równania $\nabla_1 L \cdot \nabla_0 L = 0$ wyznaczamy jako z liniowego równania pierwszego rzędu ${}_1 L$; drugie równania pary (13b), przy znanym ${}_0 A$, ${}_0 L$, ${}_1 L$, jest znowu równaniem liniowym pierwszego rzędu na ${}_1 A$.

Ogólnie biorąc, pierwsze z równań (13c) przy znanych ${}_0 A, \dots, {}_{n-1} A$, ${}_0 L, \dots, {}_{n-1} L$ stanowi równanie liniowe niejednorodne pierwszego rzędu na ${}_n L$. Wyznaczając zeń ${}_n L$, a więc znając ${}_0 A, \dots, {}_{n-1} A$, ${}_0 L, \dots, {}_{n-1} L$ drugie równanie pary (13c) można uważać za liniowe równanie niejednorodne na funkcję ${}_n A$. Jasny jest przeto ogólny mechanizm rekurencji. Jeśli więc tylko sprecyzujemy warunki brzegowe dla ${}_0 L$, ${}_1 L, \dots, {}_n L$, ${}_0 A$, ${}_1 A, \dots, {}_n A$, to procedura wyliczania

$$A_l = \sum_{n=0}^l \frac{1}{k_0^n} {}_n A, \quad L_l = \sum_{n=0}^l \frac{1}{k_0^n} {}_n L$$

stanie się jednoznacznie określoną.

Pamiętając, że nasza procedura ma na celu coraz lepsze sukcesywne zbliżanie się do całek układu (3a), (3b), przy warunkach brzegowych $A=A$ i $L = \frac{\bar{\vartheta}}{k_0}$ na Σ , postąpimy w ten sposób, iż cały ciężar warunków brzegowych przerzucimy na równania zerowego przybliżenia. Załóżmy więc, że

$$\begin{aligned} {}_0L &= \frac{\bar{\vartheta}}{k_0}, & {}_0A &= \bar{A} \\ x \in \Sigma \rightarrow {}_nL &= 0, & {}_nA &= 0 \quad n \geq 1. \end{aligned} \quad (14)$$

Zgodnie z tym co powiedziano wyżej, procedura sukcesywnego wyznaczania A_n i L_n staje się wówczas jednoznaczna.

Nie posiadamy wprowadzić matematycznego dowodu, lecz wypowiadamy intuicyjną ze względów fizycznych hipotezę, że granice z tak uzyskanych amplitud i eikonatów do rzędu L , przy dostatecznie regularnych warunkach brzegowych będą istniały i że będą one rozwiązaniem podstawowego problemu brzegowego dla równań (3a), (3b).

Interpretacja optyczna podana w rozdziale 2 pozwala teraz wykorzystać fizycznie poszczególne kroki przybliżeń. Zajmiemy się nimi po kolei.

A. Optyka do rzędu 0.

Oczywiście układ równań (13a), czyli

$$(\nabla_0 L)^2 - n^2 = 0, \quad \nabla_0(A^2 \nabla_0 L) = 0$$

jest identyczny z klasycznym układem równań optyki geometrycznej.

Znajdując przy warunkach ${}_0L = \frac{\bar{\vartheta}}{k_0}$ i ${}_0A = \bar{A}$ na Σ rozwiązanie tego układu metodą charakterystyk Cauchego, dysponujemy ${}_0A$ i ${}_0L$ a zarazem początkiem naszego ciągu przybliżeń do ścisłego rozwiązania $A_0 = {}_0A$, $L_0 = {}_0L$. Możemy więc utworzyć przybliżone rozwiązanie w rzędzie, zero jako

$$U_0 = \text{Re } A_0 \exp [i(k_0 L_0 + \omega t)].$$

Współczynnik załamania N dany jest w tym rzędzie po prostu jako $N_0 = n$, to jest redukuje się do współczynnika załamania wywołanego niejednorodnością ośrodka.

B. Optyka do rzędu 1

Przy znanych ${}_0L$ i ${}_0A$ równania (13b)

$$\nabla_1 L \cdot \nabla_0 L = 0, \quad 2\nabla_1 A \cdot \nabla_0 L + {}_1A \cdot \Delta_0 L + 2\nabla_0 A \cdot \nabla_1 L + {}_0A \Delta_1 L = 0$$

posiadają, przy zerowych warunkach brzegowych (14), wyłącznie zerowe rozwiązania ${}_1L = 0$, ${}_1A = 0$. Zatem dzięki doborowi warunków brzegowych

$$A_1 = {}_0A, \quad L_1 = {}_0L, \quad U_1 = U_0.$$

Oczywiście, $N_1 = n$. Zatem optyka do rzędu zero jest identyczna z optyką do rzędu 1.

C. Optyka do rzędu 2

Znając ${}_0A$, ${}_0L$ i ${}_1A=0$, ${}_1L=0$ z równań (13c) wnioskujemy, że ${}_2A$ i ${}_2L$ należy wyznaczyć z równań

$$\nabla_2 L \cdot \nabla_0 L = \frac{1}{2} \frac{\Delta_0 A}{{}_0A}, \quad 2\nabla_2 A \cdot \nabla_0 L + {}_2A \Delta_0 L + 2\nabla_0 A \cdot \nabla_2 L + {}_0A \Delta_2 L = 0. \quad (15)$$

Równania te przy zerowych warunkach początkowych na Σ posiadają w ogólności niezerowe jednoznacznie określone rozwiązania. Rozwiązując je i dysponując ${}_2L$, ${}_2A$ mamy wtedy również

$$A_2 = {}_0A + \frac{1}{k_0^2} {}_2A, \quad L_2 = {}_0L + \frac{1}{k_0^2} {}_2L$$

tak, że możemy napisać jako przybliżone rozwiązanie ścisłych równań do rzędu 2

$$U_2 = \text{Re } A_2 \exp [i(k_0 L_2 + \omega t)].$$

Ze względu jednak na stosowalność optycznej interpretacji do ścisłych równań, możemy również i to przybliżone rozwiązanie interpretować optycznie. W szczególności mając eikonał do rzędu 2, L_2 , linie wektorowe pola ∇L_2 interpretujemy jako promienie do rzędu 2. Można pokazać, że kwadrat amplitudy — z dokładnością do rzędu dwa propaguje się po tych liniach, to jest że

$$\nabla(A_2^2 \nabla L_2) = 0 + O\left(\frac{1}{k_0^3}\right).$$

Współczynnik załamania N z dokładnością do tego rzędu jest dany jako

$$N_2 = \sqrt{n^2 + \frac{1}{k_0^2} \frac{\Delta_0 A_2}{{}_0A}}$$

z dokładnością do $\frac{1}{k_0^2}$, to jest jako

$$N_2 = + \frac{1}{2k_0^2} \frac{\Delta_0 A}{{}_0A}.$$

D. Optyki do rzędu 1.

Nie wdając się w szczegółową dyskusję stwierdzimy tylko, iż posługując się wyżej określoną procedurą matematyczną łatwo dojdziemy do A_1 i L_1 , a więc i do U_1 , jak i N_1 ; linie wektorowe pola ∇L_1 dają nam promienie z dokładnością do tego rzędu.

Jasne jest, że — w założeniu zbieżności — im wyżej pójdziemy w rzędzie optyki, tym bliżsi będziemy rzeczywistej optyki rzędu ∞ . Jak się

wydaje, pierwsze kroki mogą dość dobrze oddawać rzeczywistość jedynie wtedy, gdy będą spełnione założenia podane w rozdziale drugim, to jest gdy $\frac{1}{k_0^2} \frac{\Delta A_l}{A_l}$ będzie małe w porównaniu z n oraz gdy A_l będzie wolno-zmienną funkcją. Czy ostatnie założenia będą spełnione, będzie oczywiście zależać w dużym stopniu od charakteru warunków brzegowych $\bar{A}$ i $\bar{\vartheta}$ na Σ .

Zrobimy jeszcze jedną uwagę. Nasza interpretacja i metoda przybliżzeń sugeruje jeszcze inną wersję *polepszenia* optyki geometrycznej. Przypuśćmy, że znamy rozwiązanie zerowego kroku ${}_0L$ i ${}_0A$ przy określonych warunkach brzegowych. Znamy jednak wtedy współczynnik załamania do rzędu 2

$$N_2 = n + \frac{1}{2k_0^2} \frac{\Delta {}_0A}{{}_0A}.$$

Otóż jest rzeczą dość naturalną intuicyjnie, aby rozwiązać teraz następujący problem

$$(\nabla \tilde{L})^2 = N_2^2, \quad \nabla (\tilde{A}^2 \nabla \tilde{L}) = 0 \quad (16)$$

przy $\tilde{L} = \bar{L} = \frac{\bar{\vartheta}}{k_0}$, $\tilde{A} = \bar{A}$ na Σ , nie czyniąc zaniedbań w rzędach $\frac{1}{k_0}$.

Jasne jest, że tak uzyskane rozwiązanie będzie z dokładnością do $O\left(\frac{1}{k_0^3}\right)$ koincydować z rozwiązaniem optyki do rzędu 2. Jeżeli jednak nie interesujemy się wyższymi rzędami, możemy poprzestać w praktyce na analitycznych wyrażeniach na $\tilde{L}$ i $\tilde{A}$. Fakt ten rzuca nowe światło na naszą interpretację i metodę.

Oczywiście, operując schematem równań (15) uwzględniamy z dokładnością do $\frac{1}{k_0^2}$ rzeczywisty mechanizm oddziaływania między fazą i amplitudą. Pozostajemy jednak w prostym kręgu pojęciowym optyki geometrycznej, tyle że z innym współczynnikiem załamania. Możemy zatem posługiwać się całym jej prostym aparatem z prawami odbicia, załamania³ itp.

Jeśli przybliżenie z N_2 okaże się niewystarczające, wówczas trzeba pójść o odpowiednią liczbę kroków dalej. Jeśli dojdziemy do rozwiązań równań optyki aż do rzędu $l-1$, jasne że będziemy dysponować N_l ; stawiając teraz, przy znanym $N_l(\mathbf{x})$, problem jako problem optyczny

$$(\nabla \tilde{L})^2 = N_l^2, \quad \nabla (\tilde{A}^2 \nabla \tilde{L}) = 0.$$

³ O tyle, o ile zagadnienie to można traktować na gruncie skalarnym.

Przy $\tilde{L} = \frac{\bar{\vartheta}}{k_0}$ i $\tilde{A} = \bar{A}$ na Σ możemy znowu wykorzystać prosty aparat optyki, mając jednocześnie pewność, iż rozwiązanie

$$\tilde{U} = \mathcal{R}_e \tilde{A} \exp [i(k_0 \tilde{L} + \omega t)]$$

różni się od rzeczywistego rozwiązania w rzędzie $O\left(\frac{1}{k_0^{l+1}}\right)$.

Jak nam się wydaje, właśnie w możliwości posługiwania się $\tilde{A}$ i $\tilde{L}$ przy znanym $N_l(\mathbf{r})$, gdzie można już zapomnieć o metodzie aproksymacji, leży główna ewentualna wartość praktyczna naszej metody.

Dorzucimy jeszcze uwagę o fizycznej stronie intuicji naszego postawienia problemu. Przypuśćmy, że na przykład ośrodek jest jednorodny — dla ustalenia uwagi jest próżnią — tak że $\varepsilon = \varepsilon_0$, $\mu = \mu_0$.

Wyliczając N_2 stwierdzimy, że $N_2 = N_2(\mathbf{x})$. Częściowe już uwzględnienie oddziaływania fazą z amplitudą przejawia się jako pewnego rodzaju indukowana niejednorodność w przestrzeni. Jeśli uznamy pojęcia amplitudy i fazy (które są wielkościami mierzalnymi) za dane i operujemy ściśle określonym pojęciem lokalnej prędkości fazowej, to niejednorodność przestrzeni, uwarunkowana przez warunki brzegowe, polegająca na fakcie, iż $N = N(\mathbf{x})$ jest równie realna i mierzalna fizycznie jak niejednorodność związana na przykład z nierównomiernym rozkładem gęstości masy w dielektryku.

4. PRZYKŁAD NA ZASTOSOWANIE OPTYKI DO RZĘDU 2

Przypuśćmy, że w próżni ($\varepsilon = \varepsilon_0$, $\mu = \mu_0$, tak że $n = 1$) na powierzchni nieskończonego walca o promieniu R , który traktujemy jako powierzchnię Σ , dane są $\bar{\vartheta} = 0$ i $\bar{A} = af(\varphi)$, gdzie φ jest kątem biegunowym. Funkcja $f(\varphi)$ spełnia warunek $f(\varphi + 2\pi) = f(\varphi)$ i jest dodatnia to jest $f > 0$. Pytamy o pole na zewnątrz walca, to jest o powierzchnie ekwifazowe oraz o rozkład amplitudy. Ze względu na symetrię problemu zarówno L jak i A nie zależą od trzeciej współrzędnej walcowej z , tak iż problem można traktować jako płaski, szukając $L = L(r, \varphi)$ i $A = A(r, \varphi)$.

Fizycznie jest jasne, że kształt powierzchni ekwifazowych powinien zależeć od rozkładu amplitudy, to jest od $f(\varphi)$.

Co zaś nam mówi optyka do rzędu zero, czyli optyka geometryczna? Równanie eikonalne zapisane we współrzędnych cylindrycznych będzie

$$\left(\frac{\partial_0 L}{\partial r}\right)^2 + \frac{1}{r^2} \left(\frac{\partial_0 L}{\partial \varphi}\right)^2 + \left(\frac{\partial_0 L}{\partial z}\right)^2 = 1$$

i jego rozwiązaniem spełniającym warunek brzegowy $_0L = 0$, gdy $r = R$, jest jednoznacznie $_0L = r - R$. Optyka geometryczna mówi więc, że po-

wierzchnie ekwifazowe są rodziną walców współosiowych. Ponieważ $\nabla_r L = i_r$, zatem promienie są rodziną prostych przechodzących przez oś z i prostopadłych do tej osi.

Równanie $\nabla({}_0A^2 \nabla_0 L) = 0$ można napisać jako

$$\nabla({}_0A^2 i_r) = \frac{1}{r} \frac{\partial}{\partial r} r_0 A^2 = 0,$$

skąd

$${}_0A = \frac{c(\varphi)}{\sqrt{r}}.$$

Pragnąc spełnić warunek brzegowy na ${}_0A$ musimy wybrać

$$c(\varphi) = \sqrt{R} a f(\varphi),$$

tak że

$${}_0A = a \sqrt{\frac{R}{r}} f(\varphi).$$

Ze względu na wspomnianą intuicję fizyczną rozwiązanie to oczywiście jest niezadowolające. Pozwala ono jednak zbudować współczynnik załamania

$$\begin{aligned} N_2 = 1 + \frac{1}{2k_0^2} \frac{\Delta_r A}{{}_c A} &= 1 + \frac{1}{2k_0^2 A} \left(\frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \varphi^2} \right) {}_0A = \\ &= 1 + \frac{1}{8k_0^2 r^2} \left(1 + 4 \frac{f''(\varphi)}{f(\varphi)} \right), \end{aligned}$$

który w ogólności zależy od r i φ . Posługując się nim można by rozwiązać układ

$$(\nabla \tilde{L})^2 - N_2^2 = 0, \quad \nabla(\tilde{A}^2 \nabla \tilde{L}) = 0$$

i uzyskać rozwiązanie koincydujące do rzędu 2 z optyką do rzędu 2, uwzględniające oddziaływanie amplitudy na fazę z dokładnością do rzędu 2. Tutaj prościej jest jednak bezpośrednio znaleźć rozwiązanie w przybliżeniu optyki do rzędu 2.

Równanie

$$\nabla_2 L \cdot \nabla_0 L = \frac{1}{2} \frac{{}_0A}{{}_0A}$$

redukuje się u nas do

$$\frac{\partial}{\partial r} {}_2L = \frac{1}{8r^2} \left(1 + 4 \frac{f''(\varphi)}{f(\varphi)} \right),$$

skąd wyliczamy ${}_2L$ spełniające zerowy warunek brzegowy jako

$${}_2L = \frac{1}{8} \left(\frac{1}{R} - \frac{1}{r} \right) \left(1 + 4 \frac{f''(\varphi)}{f(\varphi)} \right).$$

Posiadając ${}_2L$ można wyliczyć oczywiście również ${}_2A$, co jednak nie jest dla nas ciekawe.

Dysponując ${}_2L$ posiadamy eikonał optyki do rzędu 2 jako

$$L_2 = r - R + \frac{1}{8k_0^2} \left(\frac{1}{R} - \frac{1}{r} \right) \left(1 + 4 \frac{f''(\varphi)}{f(\varphi)} \right),$$

a więc i powierzchnie ekwifazowe, których kształt uwarunkowany jest rozkładem amplitudy. Pozwala to wyliczyć linie wektorowe ∇L_2 , to jest promienie w optyce do rzędu 2. Spełniają one równanie różniczkowe

$$0 = d\mathbf{r} \times \nabla L_2 = (\mathbf{i}_r dr + \mathbf{i}_\varphi r d\varphi) \times \left[\mathbf{i}_r \left\{ \frac{1}{8k_0^2 r^2} \left(1 + 4 \frac{f''}{f} \right) + 1 \right\} + \right. \\ \left. + \mathbf{i}_\varphi \frac{1}{r} \frac{1}{8k_0^2} \left(\frac{1}{R} - \frac{1}{r} \right) 4 \frac{\partial}{\partial \varphi} \frac{f''}{f} \right].$$

Równanie to sprowadza się do związku

$$\frac{dr}{r} \frac{1}{8k_0^2} \left(\frac{1}{R} - \frac{1}{r} \right) 4 \frac{\partial}{\partial \varphi} \frac{f''}{f} - r d\varphi \left(1 + \frac{1}{8k_0^2 r^2} \left(1 + 4 \frac{f''}{f} \right) \right) = 0.$$

Oczywiście, linie spełniające to równanie interesują nas jedynie w przybliżeniu. Zanedbując

$$\frac{1}{8k_0^2 r^2} \left(1 + 4 \frac{f''}{f} \right)$$

wobec 1, to jest przyjmując, że

$$1 + \frac{1}{8k_0^2 r^2} \left(1 + 4 \frac{f''}{f} \right) \approx 1$$

i rozdzielając zmienne w otrzymanym po tym uproszczeniu równaniu mamy

$$\frac{1}{2k_0^2} \frac{dr}{r^2} \left(\frac{1}{R} - \frac{1}{r} \right) = \frac{d\varphi}{\frac{d}{d\varphi} \frac{f''}{f}},$$

skąd wynika, że nasze promienie dane są przez równania

$$-\frac{1}{2k_0^2} \left[\frac{1}{Rr} - \frac{1}{2} \frac{1}{r^2} \right] = \int \frac{d\varphi}{\frac{d}{d\varphi} \frac{f''}{f}}.$$

Linie te w ogólności nie są oczywiście liniami prostymi.

Dla konkretności przyjmijmy na przykład, że $f = 1 + \cos \varphi$. Wyliczenie całki daje, iż

$$-\frac{1}{2k_0^2} \left[\frac{1}{Rr} - \frac{1}{2} \frac{1}{r^2} \right] = 4 \ln \sin \frac{\varphi}{2} - 2 \sin^2 \frac{\varphi}{2} + C.$$

Rozwiązując ten związek jako równanie kwadratowe względem $\frac{1}{r}$ znajdujemy explicite równanie biegunowe jednoparametrowej rodziny promieni

$$r = r(\varphi, C).$$

5. PRZYBLIŻENIA METODY W.K.B.

Jak podano w rozdziale drugim zastosowanie metody W.K.B. zmienia problem całkowania równania Helmholtza

$$\Delta \varphi + k_0^2 n^2 \varphi = 0 \quad (17)$$

z warunkiem brzegowym

$$\varphi|_{\Sigma} = A \exp i\vartheta \quad (18)$$

na problem całkowania układu równań

$$(\nabla_0 A)^2 = n^2, \quad (19_0)$$

$$\dots, \dots$$

$$2\nabla_n A \cdot \nabla_0 A = -\Delta_{n-1} A - \sum_{p=1}^{n-1} \nabla_p A \cdot \nabla_{n-p} A, \quad \text{dla } n \geq 1 \quad (19_n)$$

przy warunkach początkowych

$${}_0 A|_{\Sigma} = \frac{\vartheta}{k_0} + \frac{1}{ik_0} \ln A; \quad n A|_{\Sigma} = 0 \quad \text{dla } n \geq 1. \quad (20)$$

Warunek (18) i warunki (20) są zadane na tej samej powierzchni (lub krzywej w przypadku dwuwymiarowym). Jaki charakter mają te warunki można dopiero stwierdzić biorąc pod uwagę równanie, do którego one się odnoszą. Nakłada to, niestety, pewne ograniczenia na klasę problemów, które podlegają omawianej teraz transformacji. Transformacja taka wydaje się na przykład niemożliwa w przypadku obszaru wewnętrznego względem powierzchni Σ . Odpowiada to intuicyjnemu obrazowi zjawisk optycznych, jakie powinny w takim obszarze wystąpić, a mianowicie interferencji i odbić. Fizyczny obraz równania (19_n) z warunkami (20) odpowiadający problemowi przetransformowanemu, jest to propagacja pewnych powierzchni ekwifazowych. O ile powierzchnie te trafiają na siebie, następuje interferencja, a wynik jej nie stanowi rozwiązania problemów (19_n) i (20).

Można by próbować inaczej rozbić warunki graniczne pomiędzy poszczególne rzędy, niż to zostało zrobione w układzie (20). Skoncentrowanie jednak ich w rzędzie zerowym pozwala mieć nadzieję, że już w zerowym kroku aproksymacji uda się uchwycić istotne cechy rozwiązania. Czy

poprzestając na zerowym kroku aproksymacji jesteśmy w posiadaniu bogatszego opisu procesu rządzonego przez równanie Helmholtza niż w klasycznej optyce geometrycznej? Niewątpliwie tak.

Równanie (19₀) pozornie o postaci identycznej z równaniem $(\nabla L)^2 = 1$, różni się od niego w sposób istotny, mianowicie ${}_0A$ jest zespolone. Wprowadzając oznaczenia dla części rzeczywistej i urojonej

$${}_0A = u + iv, \quad (21)$$

z równania (19₀) otrzymujemy następujące zależności:

$$(\nabla u)^2 - (\nabla v)^2 = n^2, \quad (22)$$

$$\nabla u \cdot \nabla v = 0. \quad (23)$$

Z powyższego wynika, że zerowe przybliżenie będzie miało postać

$$\varphi_0 = \exp ik_0 A_0 = \exp(-k_0 v) \cdot \exp ik_0 u. \quad (24)$$

Zatem już zerowe przybliżenie dopuszcza zmiany amplitudy wyrażone przez czynnik $\exp(-k_0 v)$. Ponadto równania (22) i (23) sprzęgają ze sobą zmiany amplitudy i fazy. Z równania (22) wynika, że zmiany amplitudy pociągają za sobą wzrost $|\nabla u|$, a co za tym idzie, zmniejszenie prędkości fazowej. Elementarnym przykładem, w którym zjawisko to może wystąpić i występuje, jest fala płaska niejednorodna. Fakt, że sprzężenie to występuje już w zerowym kroku aproksymacji, pozwala przypuszczać, iż metoda W.K.B. okaże się korzystna w zastosowaniu do zagadnień mikrofalowych, gdzie niejednokrotnie zależy na uchwyceniu wpływu sprzężenia amplitudy z fazą.

Przybliżenia metody W.K.B. w zasadniczo inny sposób segregują na rzędy wielkości względem $\frac{1}{k_0}$ rozwiązanie φ , niż to czyniła metoda omawiana w rozdziałach poprzednich. Pisząc

$$\varphi = \exp \left[-k_0 \sum_{n=0}^{\infty} \mathcal{I}_n \left(\frac{1}{ik_0} \right)^n {}_nA \right] \cdot \exp \left[ik_0 \sum_{n=0}^{\infty} \mathcal{R}_n \left(\frac{1}{ik_0} \right)^n {}_nA \right] \quad (25)$$

widzimy, że w metodzie W.K.B. faza rozwija się wprawdzie tak jak w metodzie poprzedniej na szereg potęgowy względem $\frac{1}{k_0}$, amplituda jednak rozwija się na iloczyn czynników o postaci widocznej z powyższego wzoru.

Łatwo również zauważyć, że zerowe przybliżenie omawianej obecnie metody jest w podobnym stosunku do fal płaskich niejednorodnych, jak zerowe przybliżenie poprzedniej metody do zwyczajnych fal płaskich.

Jak wiadomo, równanie

$$\Delta \varphi + k_0^2 \varphi = 0 \quad (26)$$

posiada rozwiązania typu

$$\varphi = \exp [ik_0(ax + by + cz + d)],$$

gdzie a, b, c, d są stałymi zespolonymi spełniającymi warunek

$$a^2 + b^2 + c^2 = 1. \quad (27)$$

Rozwiązanie takie nazywa się falą płaską niejednorodną.

Zbadajmy teraz zerowe przybliżenie według metody W.K.B. do rozwiązania równania (26). Równanie to powstaje z równania (17) przez założenie, że $n=1$ w interesującym nas obszarze. Rozwijamy wykładnik wyrażenia

$$\varphi_0 = \exp ik_0 A_0 \quad (28)$$

w otoczeniu ustalonego punktu $\mathbf{x}_0$ i zanedbujemy małe wyższych rzędów względem $\mathbf{x} - \mathbf{x}_0$; daje to

$$\varphi = \exp \{ ik_0 [\mathbf{x} \nabla A_0(\mathbf{x}_0) - \mathbf{x}_0 \nabla A_0(\mathbf{x}_0) + A_0(\mathbf{x}_0)] \}. \quad (29)$$

Otrzymane wyrażenie jest falą płaską niejednorodną o parametrach

$$a = \left(\frac{\partial A_0}{\partial x} \right)_{\mathbf{x}_0}, \quad b = \left(\frac{\partial A_0}{\partial y} \right)_{\mathbf{x}_0}, \quad c = \left(\frac{\partial A_0}{\partial z} \right)_{\mathbf{x}_0}, \quad d = -\mathbf{x}_0 \nabla A_0(\mathbf{x}_0) + A_0(\mathbf{x}_0). \quad (30)$$

Łatwo zauważyć, że zachodzi związek (27) charakterystyczny dla fal płaskich niejednorodnych; jest on mianowicie równoważny równaniu różniczkowemu dla zerowego przybliżenia (19₀). Tę własność pól opisywanych przez zerowe przybliżenie według metody W.K.B. wyrażamy mówiąc, że zachowują się one lokalnie jak fala płaska niejednorodna.

Przy poszukiwaniu związku zwyczajnych fal płaskich z zerowym przybliżeniem poprzedniej metody występuje jeszcze jeden dodatkowy moment. Jeżeli odpowiednie wyrażenie przekształcimy jako

$$\varphi = A_0 \exp [ik_0 L_0] = \exp \left[ik_0 \left(L_0 + \frac{1}{ik_0} \ln A_0 \right) \right], \quad (31)$$

a następnie wykładnik rozwinieśmy wokół ustalonego punktu $\mathbf{x}_0$ na szereg potęgowy, to aby wyrażenie to uprościło się do fali płaskiej należy nie tylko zaniebgać wyrazy rzędu drugiego i wyższych względem $\mathbf{x} - \mathbf{x}_0$,

lecz również $\frac{1}{ik_0} \frac{\nabla A_0}{A_0}$ w porównaniu do ∇L_0 . Innymi słowy zakłada się,

że zmiany amplitudy na odcinkach rzędu długości fali — w odróżnieniu od zmian fazy — są zanedbywalne. Sprawdzenie, że przy omówionych uproszczeniach wyrażenie redukuje się do fali płaskiej, jest trywialne. Założenie zrobione przy porównywaniu zerowych przybliżeń obu metod z falami płaskimi, że $n=1$ w rozważanym obszarze, można zastąpić założeniem, że rozpatrujemy te zerowe przybliżenia w obszarze o n mało zmieniającym się na odcinkach rzędu długości fali.

Wprawdzie przekonywającego porównania obu metod aproksymacji można by się spodziewać dopiero po szerszym ich zastosowaniu do konkretnych zagadnień, wydaje się jednak, że metoda druga może się okazać lepiej dostosowana do badania zagadnień, w których zmiany amplitudy na odcinkach rzędu długości fali nie są zaniedbywalne, ze względu na to, że składnik wykładnika, od którego zależą zmiany amplitudy, posiada wyraz pierwszego rzędu względem k_0 .

Ze względu na to, że równania, do których doprowadziła metoda W.K.B., nie są objęte klasycznymi twierdzeniami z dziedziny równań różniczkowych cząstkowych, należy jeszcze zweryfikować poprawność w ten sposób postawionego problemu matematycznego. W dotychczasowych rozważaniach, ze względów poglądowych, rozważano na ogół ośrodek o zmiennym współczynniku załamania n . Pozwoliło to, między innymi, na wykazanie, że odchylenia, jakie wykazują ściśle rozwiązania równań elektrodynamiki w stosunku do rozwiązań równań optyki, dadzą się sprowadzić do pewnej modyfikacji współczynnika załamania ośrodka i rozważania pól o charakterze optyki geometrycznej w ośrodku o zastępczym współczynniku załamania.

Dla zastosowań praktycznych, które niniejsza praca ma na widoku, a mianowicie dla zagadnień optyki mikrofalowej, zasadnicze znaczenie ma przypadek ośrodków o stałym współczynniku załamania. Ograniczenie się do takich ośrodków wprowadzi do rozważań matematycznych zasadnicze uproszczenia. Nieistotne z punktu widzenia ogólności rozważań (dające się łatwo usunąć) będzie założenie, że $n=1$.

Rozważmy obecnie problem Cauchy dla równania

$$(\nabla_0 A)^2 = 1. \quad (32)$$

Zajmiemy się problemem płaskim, to znaczy przypadkiem, gdy ${}_0A$ jest funkcją dwóch zmiennych. Znaczna część rezultatów daje się bez trudu przenieść na przypadek trzech zmiennych.

Dana jest krzywa bez punktów podwójnych. Jako parametr służy długość łuku $x=x_0(r)$, $y=y_0(r)$, $\dot{x}_0^2 + \dot{y}_0^2 = 1$, $\alpha \leq r \leq \beta$. (33)

Na krzywej tej określone są wartości funkcji ${}_0A = u + iv$

$$u = u_0(r), \quad v = v_0(r), \quad l_0 = u_0 + iv_0. \quad (34)$$

O funkcjach x_0, y_0, u_0, v_0 , zmiennej rzeczywistej r zakładamy na razie, że względem zmiennej zespolonej $t = r + is$ są one regularne w pewnym zbiorze otwartym, zawierającym odcinek osi rzeczywistej $\alpha < r \leq \beta$ ⁴. Rozsze-

⁴ Oznacza to założenie, że funkcje określone są za pomocą wyrażeń pozwalających na podstawienie zespolonych wartości argumentu, które to wyrażenia w wymienionym wyżej obszarze określają jednoznacznie funkcję zmiennej zespolonej posiadającą pochodną.

rzamy na dziedzinę zespoloną pojęcie całki zupełnej. Dla równania (32) będzie ona miała postać

$${}_0A = ax + by + c, \quad (35)$$

gdzie a , b i c są to liczby zespolone spełniające związek

$$a^2 + b^2 = 1. \quad (36)$$

Wyznaczamy teraz trzy funkcje analityczne $a(t)$, $b(t)$, $c(t)$ z układu równań:

$$u_0 + iv_0 = ax_0 + by_0 + c, \quad (37)$$

$$u'_0 + iv'_0 = ax'_0 + by'_0, \quad (38)$$

$$a^2 + b^2 = 1. \quad (39)$$

Jest to efektywnie wykonalne i prowadzi do wyrażen:

$$a(t) = l'_0 x'_0 + y'_0 \sqrt{1 - (l'_0)^2}, \quad (40)$$

$$b(t) = l'_0 y'_0 - x'_0 \sqrt{1 - (l'_0)^2}, \quad (41)$$

$$c(t) = l_0 - ax_0 - by_0. \quad (42)$$

Aby te trzy funkcje były w otoczeniu odcinka osi r , $\alpha \leq r \leq \beta$ regularne, wystarczy założyć, że dla tych wartości $(l'_0)^2 \neq 1$. Do tego warunku jeszcze wrócimy.

Rozwiązanie problemu Cauchy otrzymuje się za pomocą układu równań

$${}_0A = a(t)x + b(t)y + c(t), \quad (43)$$

$$0 = a'(t)x + b'(t)y + c'(t). \quad (44)$$

Należy mianowicie z tego układu równań wyrugować zmienną t . Otrzyma się wtedy związek między x , y i ${}_0A$, który wartościom rzeczywistym x i y przyporządkowuje zespoloną wartość ${}_0A$.

W przypadku dwu zmiennych wydaje się bardziej naturalna inna droga. Z równania (44) wyznaczamy x i y jako funkcje parametrów r i s , to znaczy części rzeczywistej i urojonej zmiennej t . W tym celu wprowadzamy oznaczenia:

$$a = u_1 + iv_1, \quad b = u_2 + iv_2, \quad c = u_3 + iv_3, \quad (45)$$

$$a' = u_1^{(1)} + iv_1^{(1)}, \quad b' = u_2^{(1)} + iv_2^{(1)}, \quad c' = u_3^{(1)} + iv_3^{(1)}, \quad (46)$$

gdzie funkcje u i v z różnymi indeksami są rzeczywiste. x i y są zatem wyznaczone przez układ dwóch równań algebraicznych liniowych

$$\begin{aligned} u_1^{(1)}x + u_2^{(1)}y + u_3^{(1)} &= 0, \\ v_1^{(1)}x + v_2^{(1)}y + v_3^{(1)} &= 0. \end{aligned} \quad (47)$$

Zajmiemy się teraz wyznacznikiem tego układu równań. O ile a' i b' znikają w rozważanym obszarze, to z równania (44) wynika, że w obszarze tym znika i c' , a stąd wynika, że ${}_0A$ redukuje się do postaci (35) i żadne rugowanie parametru nie jest potrzebne. O ile jedna z funkcji a' i b' jest stale różna od zera, powiedzmy że jest to a' , wtedy $\frac{b'}{a'}$ jest rzeczywiste, gdy tylko wyznacznik układu znika. Różniczkujemy teraz równanie (39), co daje

$$\frac{b'}{a'} = -\frac{a}{b}, \quad (48)$$

zatem wtedy i stosunek $\frac{a}{b}$ jest rzeczywisty, co pozwala łatwo z równania (39) wyprowadzić wniosek, że a i b są rzeczywiste i wówczas problem redukuje się do optyki geometrycznej. Zatem w przypadku, gdy mamy do czynienia z optyką klasyczną, postępowanie takie nie jest możliwe.

Po wyznaczeniu x i y z układu równań (47) jako funkcji dwóch parametrów, równanie (43) daje parametryczne równanie zespolonej powierzchni całkowitej spełniającej postawiony problem początkowy. Ciekawy jest fakt, że wskazane postępowanie, prowadzące do otrzymania parametrycznego równania powierzchni całkowitej, jest efektywnie wykonalne do końca, jeżeli tylko potrafimy rozdzielić funkcje a' , b' i c' na części rzeczywiste i urojone. Jest to dość rzadki w dziedzinie równań cząstkowych przykład efektywnie otrzymywalnego rozwiązania.

Ciekawy wydaje się fakt, że charakter analitycznych zależności daje się tylko częściowo narzucić na postawiony problem. Mianowicie z równań (22) i (23) wynika, że funkcja ${}_0A$, traktowana jako funkcja zmiennej zespolonej $z=x+iy$, nie może być analityczna, gdyż wymienione równania są sprzeczne z równaniami Cauchy—Riemanna. Również można udowodnić, że odwzorowanie płaszczyzny $z=x+iy$ na płaszczyznę $t=r+is$ nie ma charakteru analitycznego. Nie może być również analityczna zależność ${}_0A$ od t .⁵

Pozostało jeszcze do omówienia założenie $(l'_0)^2 \neq 1$ zrobione w trakcie dowodu. Równanie $(l'_0)^2 = 1$ jest równoważne spełnieniu w danym punkcie krzywej dwóch zależności

$$\nabla v = 0, \quad u_x \dot{x}_0 + u_y \dot{y}_0 = 1, \quad (49)$$

które oznaczają, że w danym punkcie krzywej nie tylko jest stacjonarna amplituda, lecz również kierunek krzywej początkowej jest zgodny z kierunkiem promienia (promienie — linie pola ∇u). Poczynione założenie eliminuje zatem charakterystyczny kierunek krzywej początkowej, po-

⁵ To znaczy, że nie można poszukiwać tych zależności w postaci funkcji posiadających pochodną w dziedzinie zespolonej.

dobnie jak to należy zrobić dla równań w dziedzinie rzeczywistej. W przypadku, gdy dane początkowe znajdują się na krzywej ekwifazowej, założenie to jest automatycznie spełnione.

Oczywiście, otrzymane rozwiązanie problemu Cauchy ma charakter lokalny, zapewnia bowiem istnienie tego rozwiązania w pewnym otoczeniu krzywej początkowej. Wystąpienie punktów osobliwych rozważanych funkcji, poza krzywą początkową, jest możliwe choćby w przypadku, gdy dwa różne fragmenty fali docierają do tego samego punktu.

Podstawową rolę dla równań różniczkowych o pochodnych cząstkowych pierwszego rzędu gra pojęcie charakterystyk. Rozważane jednak równanie odnosi się do funkcji zespolonych i, jak widzieliśmy, jest ono w dziedzinie rzeczywistej równoważne układowi dwóch równań (22) i (23). Z tego względu o charakterystykach w normalnym znaczeniu trudno tu właściwie mówić. Istotne znaczenie z punktu widzenia zastosowań posiadają natomiast linie pola ∇u (promienie) i ∇v . Linie te są blisko spokrewnione z rzutami charakterystyk równań w dziedzinie rzeczywistej.

Przechodzimy obecnie do wyznaczenia równań tych krzywych. Według oznaczeń (45) wektory ∇u i ∇v posiadają następujące współrzędne

$$\nabla u = [u_1, u_2], \quad \nabla v = [v_1, v_2]. \quad (50)$$

Jeżeli a i b wyznaczyliśmy jako funkcje x i y , równanie różniczkowe określające linie pola ∇u przyjmie postać

$$y' = \frac{u_2(x, y)}{u_1(x, y)} \quad (51)$$

i analogiczne równanie otrzymać można dla linii pola ∇v .

a i b , a co za tym idzie u_1, u_2, v_1, v_2 , na ogół biorąc nie dają się jednak wyrazić *explicite* jako funkcje x i y . Odpowiednie równania różniczkowe dają się otrzymać w postaci wyrażonej w przypadku, gdy będziemy się posługiwali poprzednio wprowadzonymi parametrami r i s . Zagadnienie znalezienia krzywych, o których mowa, sprowadza się oczywiście do wyznaczenia odpowiedniego związku między r i s , powiedzmy w postaci funkcji $r = f(s)$. Podstawiając do rozwiązań układu równań (47) $x(r, s)$ i $y(r, s)$ na $r, f(s)$ otrzymujemy parametryczne równanie krzywej.

Dzięki zależności

$$aa' + bb' = 0 \quad (52)$$

otrzymanej przez różniczkowanie równania (39) otrzymuje się z równania (44), mnożąc je przez $-\frac{a}{b'}$, zależność

$$bx - ay - \frac{c'a}{b'} = 0. \quad (53)$$

Wprowadzamy następujące oznaczenia dla części rzeczywistej i urojonej

$$-\frac{c'a}{b'} = u_4 + iv_4. \quad (54)$$

Wtedy równanie (53) jest równoważne następującemu układowi równań rzeczywistych:

$$u_2x - u_1y + u_4 = 0, \quad (55)$$

$$v_2x - v_1y + v_4 = 0. \quad (56)$$

Równania te są przez $x(r, s)$ i $y(r, s)$ tożsamościowo spełniane. Niech $r = f(s)$ określa w wyżej podany sposób linię pola ∇u . Podstawiamy w tożsamości (55) $r = f(s)$ i różniczkujemy względem s . Dzięki зробionemu założeniu odnośnie funkcji f część składników odpadnie i otrzymamy

$$(u_{2r}f' + u_{2s})x - (u_{1r}f' + u_{1s})y + u_{4r}f' + u_{4s} = 0. \quad (57)$$

Ze względu na związki Cauchy—Riemanna i równania (45) i (46), otrzymane wyrażenie można napisać nieco inaczej jako

$$(u_2^{(1)}f' - v_2^{(1)})x - (u_1^{(1)}f' - v_1^{(1)})y + u_{4r}f' + u_{4s} = 0. \quad (58)$$

Ostatecznie, ze względu na układ równań (47) oraz na równanie (58), otrzymujemy następujące równanie różniczkowe określające rodzinę linii pola ∇u

$$\begin{vmatrix} u_1^{(1)} & u_2^{(1)} & u_3^{(1)} \\ v_1^{(1)} & v_2^{(1)} & v_3^{(1)} \\ \dots & \dots & \dots \\ (u_2^{(1)}f' - v_2^{(1)}) & (-u_1^{(1)}f' - v_1^{(1)}) & (u_{4r}f' + u_{4s}) \end{vmatrix} = 0. \quad (59)$$

Ustalając odpowiednio wartość $r = f(0)$ otrzymujemy linię pola przechodzącą przez pożądaną punkt krzywej początkowej.

Korzystając z równania (56) możemy w podobny sposób ustalić równanie różniczkowe rodziny linii pola ∇v . Zamiast korzystać z równania (53), można było rozwiązać układ równań (47). Pozostaje kwestią otwartą, która z wymienionych dróg:

- a) korzystająca z równania (51),
- b) korzystająca z równania (59),
- c) ostatnio wspomniana,

będzie w konkretnych przypadkach bardziej efektywna.

W odróżnieniu od optyki geometrycznej, promienie w ośrodku jednorodnym mogą nie być proste: ich kształt zależny jest od zmian amplitudy.

Łatwo sprawdzić, że w przypadku założenia warunków początkowych odpowiadających fali płaskiej niejednorodnej metoda W. K. B. już w zero-

wym przybliżeniu prowadzi do ścisłego rozwiązania. Na podstawie dyskusji problemu Cauchy jest to również intuicyjnie jasne.

Otrzymaliśmy zerowe przybliżenie w postaci

$${}_0A = ax + by + c, \quad (60)$$

gdzie a , b i c są funkcjami x i y lub też wszystkie funkcje a , b , c , x i y są funkcjami dwóch parametrów r i s . Oczywiście, dzięki równaniu (44), pochodne cząstkowe ${}_0A$, czyli składowe $\nabla_0 A$, są odpowiednio równe

$$\frac{\partial {}_0A}{\partial x} = a, \quad \frac{\partial {}_0A}{\partial y} = b. \quad (61)$$

Przejdźmy obecnie do następnego, to znaczy pierwszego przybliżenia metody W.K.B. Przybliżenie to określone jest przez równanie

$$2\nabla_0 A \cdot \nabla_1 A = -\nabla^2 {}_0A \quad (62)$$

z warunkiem początkowym

$${}_1A \Big|_{\Sigma} = 0. \quad (63)$$

Wprowadzamy nową funkcję za pomocą podstawienia

$${}_1A = \frac{1}{2} \ln p \quad (64)$$

wtedy

$$\nabla(p\nabla_0 A) = 0. \quad (65)$$

$\nabla_0 A$ jest, jak wiemy, znanym wektorem. Pomimo szczególnie prostej postaci równania (65), autorzy znaleźli jego ogólne rozwiązanie jedynie w przypadku dwuwymiarowym. Jest nim

$$p = \nabla^2 {}_0A f(a, b), \quad (66)$$

gdzie f jest funkcją dowolną byle dostatecznie regularną.

Aby sprawdzić, że równanie (65) jest spełnione przez wyrażenie (66), należy uwzględnić zależności, jakie otrzymuje się przez różniczkowanie równania (32). W notacji tensorowej mają one postać

$${}_0A_{,kl} {}_0A_{,k} = 0, \quad (67)$$

$${}_0A_{,klm} {}_0A_{,k} + {}_0A_{,k l 0} {}_0A_{,km} = 0. \quad (68)$$

Z układu równań (67), w przypadku dwóch zmiennych, wynika

$${}_0A_{xx} \cdot {}_0A_{yy} = ({}_0A_{xy})^2, \quad (69)$$

gdzie indeksy z prawej strony oznaczają, jak zwykle, odpowiednie pochodne cząstkowe. Spełnianie równania (65) przez wyrażenie (66) sprawdzamy teraz bezpośrednio przez podstawienie uwzględniając równania od (67) do (69).

Warunek początkowy dla funkcji p , jak wynika z równań (63) i (64), ma postać

$$p|_x=1. \quad (70)$$

Aby wykazać, że wyrażenie (66) jest rzeczywiście rozwiązaniem ogólnym należy sprawdzić, że rozwiązuje ono problem Cauchy (65) i (70). Przed tym jednak zauważymy, że w przypadku, gdy w rozważanym obszarze $\nabla^2_0 A$ znika, rozwiązaniem pierwotnego problemu początkowego (62), (63) jest $_1A \equiv 0$. Ograniczymy się tylko do przypadku, gdy w całym rozważanym obszarze $\nabla^2_0 A \neq 0$. Należy teraz wykazać, że można tak dobrać funkcję $f(a, b)$, aby na krzywej Σ zachodziło równanie

$$1 = f(a, b) \nabla^2_0 A. \quad (71)$$

Funkcja f zależy od dwóch argumentów a i b , to znaczy od pochodnych cząstkowych $_0A$ (równania 61). Właściwie jest to funkcja jednej zmiennej niezależnej; ze względu na związek między pochodnymi cząstkowymi $_0A$ (równanie 32) można by napisać $f(a, \sqrt{1-a^2})$, czego nie zrobiono ze względu na symetrię.

Jeżeli wykazemy, że w każdym punkcie krzywej początkowej co najmniej jedna z pochodnych względem długości łuku $\frac{da}{dr}$ lub $\frac{db}{dr}$ jest różna od zera, wtedy przez odpowiedni dobór funkcji $f(a, b)$ będzie można tak skompensować zmiany $\nabla^2_0 A$, aby ich iloczyn $f \nabla^2_0 A$ był stały i równy jedności, lecz

$$\frac{da}{dr} = a_x \dot{x}_0 + a_y \dot{y}_0 = {}_0A_{xx} \dot{x}_0 + {}_0A_{xy} \dot{y}_0 = \nabla^0_0 A b (-a \dot{y}_0 + b \dot{x}_0), \quad (72)$$

$$\frac{db}{dr} = b_x \dot{x}_0 + b_y \dot{y}_0 = {}_0A_{xy} \dot{x}_0 + {}_0A_{yy} \dot{y}_0 = \nabla^2_0 A a (-a \dot{y}_0 + b \dot{x}_0) \quad (73)$$

zgodnie z założeniem $\nabla^2_0 A \neq 0$. Ze względu na równanie (39) przynajmniej jedna z pochodnych cząstkowych a lub b jest różna od zera. Pozostało zatem rozważyć wyrażenie $(-a \dot{y}_0 + b \dot{x}_0)$.

Ostateczną formę wyrażeń (72) i (73) można otrzymać uwzględniając zależności:

$${}_0A_{xx} = b^2 \nabla^2_0 A, \quad (74)$$

$${}_0A_{xy} = -ab \nabla^2_0 A, \quad (75)$$

$${}_0A_{yy} = a^2 \nabla^2_0 A. \quad (76)$$

Zależności te z kolei wynikają z równania (32) oraz układu równań (67). Równanie

$$b \dot{x}_0 - a \dot{y}_0 = 0 \quad (77)$$

okazuje się bardzo interesujące. Przy uwzględnieniu równania (39) wynika

z niego, że wektor o współrzędnych a i b czyli $\nabla_0 A$ powinien być rzeczywisty i równoległy do wektora stycznego krzywej x_0, y_0 . Stąd wynika, że pochodna ${}_0A$ względem długości łuku równa jest jedności. Przypominamy, że warunek, aby pochodna ta była różna od jedności, wystąpił już poprzednio przy badaniu nieliniowego równania różniczkowego zerowego kroku przybliżenia

$$(\nabla_0 A)^2 = 1; \quad (32)$$

obecnie ten sam warunek zapewnia rozwiązalność problemu Cauchy również zespolonego równania różniczkowego, lecz tym razem liniowego

$$\nabla p \nabla_0 A + p \nabla^2 {}_0A = 0. \quad (65)$$

Warunek ten będzie się powtarzał, jak łatwo spostrzec, we wszystkich następnych krokach przybliżenia, ze względu na taką samą budowę wyrażenia zawierającego funkcję niewiadomą w odpowiednich równaniach. Należy przy tym dodać, że rozwiązania ogólne odpowiednich równań dadzą się wyrazić w postaci sumy (lub iloczynu po podstawieniu analogicznym do (64) rozwiązania szczególnego i dowolnej funkcji od $\nabla_0 A$.

To, że warunek eliminujący kierunek charakterystyczny krzywej początkowej ma taką samą postać dla równania (32) i równań typu (65'), byłoby w dziedzinie rzeczywistej oczywiste.

6. PRZYKŁAD ZASTOSOWANIA METODY W. K. B.

Wydaje się właściwe wybranie takiego przykładu, aby otrzymane wyniki można było prosto zweryfikować na innej, znanej drodze. Oczywiście pozytywny wynik takiego zastosowania wykaże jedynie, że istnieją przypadki, gdy przedstawiona metoda prowadzi do rozsądnych rezultatów, nie wykaże natomiast czy metoda taka jest potrzebna (ponieważ zgodnie z postulowanymi warunkami rezultaty te dają się prosto otrzymać na innej drodze). Sądzimy jednak, że rzeczowej oceny metody o szerokiej potencjalnej dziedzinie zastosowań, oceny określającej przydatność takiej metody, nie można oprzeć na jednym czy dwóch ad hoc skonstruowanych przykładach. Można pomyśleć wiele różnych wariantów zastosowań tej metody czy jej kombinacji z innymi metodami; można by te warianty tak dobierać, aby efektywność przedstawionej metody okazywała się możliwie duża, lecz chyba słuszne będzie odłożyć ocenę do czasu, gdy spróbuje się przez jakiś czas tą metodą praktycznie pracować.

Jak wiadomo, elementarne rozwiązanie równania Helmholtza

$$\nabla^2 \varphi + k_0^2 \varphi = 0 \quad (78)$$

otrzymane metodą rozdzielania zmiennych we współrzędnych biegunowych r, ϑ (przypadek dwuwymiarowy) ma postać

$$\varphi = A H_n^{(1)}(k_0 r) \exp[in\vartheta]. \quad (79)$$

Na obwodzie koła o promieniu $r=R$ funkcja ta przyjmuje wartości

$$\varphi_R = AH_n^{(1)}(k_0 R) \exp[in\vartheta]. \quad (80)$$

Spróbujmy teraz, czy i z jaką dokładnością rozwiązanie to zrekonstruuje metoda W. K. B.

Przyjmujemy zatem jako krzywą początkową Σ obwód koła o promieniu R . Równanie (80) określa nam warunki początkowe; należy tylko wziąć pod uwagę, że

$${}_0A \Big|_{\Sigma} = \frac{1}{ik_0} \ln \varphi_R. \quad (81)$$

Dla uproszczenia wprowadzamy oznaczenie

$$\frac{1}{ik_0} \ln AH_n^{(1)}(k_0 R) = \alpha \quad (82)$$

i nasz problem początkowy przyjmuje postać

$$x_0 = R \cos \tau, \quad y_0 = R \sin \tau, \quad (83)$$

$${}_0A \Big|_{\Sigma} = \frac{n}{k_0} \tau + \alpha. \quad (84)$$

Parametr, ze względu na przyjęte oznaczenia współrzędnych, został tu oznaczony dla uniknięcia nieporozumień przez τ , a nie jak dotychczas przez r (rozdz. 5).

Należy teraz wyznaczyć funkcje $a(t)$, $b(t)$, $c(t)$ z układu równań

$$\frac{n}{k_0} t + \alpha = aR \cos t + bR \sin t + c, \quad (85)$$

$$\frac{n}{k_0} = -aR \sin t + bR \cos t, \quad (86)$$

$$a^2 + b^2 = 1. \quad (87)$$

Wprowadzamy nową zmienną ψ , tak aby

$$a = \cos \psi, \quad (88)$$

$$b = \sin \psi. \quad (89)$$

Jest to możliwe ze względu na równanie (87), które jest tym samym spełnione. Równanie (86) daje

$$\psi = t + \arcsin \frac{n}{k_0 R}, \quad (90)$$

a równanie (85)

$$c = \frac{n}{k_0} t + \alpha - R \cos(\psi - t). \quad (91)$$

Obecnie, gdy już wyznaczone zostały funkcje a , b , c , rozwiązanie problemu otrzymamy przez rugowanie zmiennej t z układu równań

$${}_0A = \cos\left(t + \arcsin \frac{n}{k_0 R}\right) x + \sin\left(t + \arcsin \frac{n}{k_0 R}\right) y + \frac{n}{k_0} t + \alpha - R \cos\left(\arcsin \frac{n}{k_0 R}\right), \quad (92)$$

$$0 = -\sin\left(t + \arcsin \frac{n}{k_0 R}\right) x + \cos\left(t + \arcsin \frac{n}{k_0 R}\right) y + \frac{n}{k_0}. \quad (93)$$

Równania te bardzo się upraszczają przez wprowadzenie współrzędnych biegunowych, a mianowicie

$${}_0A = r \cos\left(\vartheta - t - \arcsin \frac{n}{k_0 R}\right) + \frac{n}{k_0} t + \alpha - R \cos\left(\arcsin \frac{n}{k_0 R}\right), \quad (92')$$

$$0 = r \sin\left(\vartheta - t - \arcsin \frac{n}{k_0 R}\right) + \frac{n}{k_0}. \quad (93')$$

Łatwo teraz wyrugować t , co daje

$${}_0A = r \sqrt{1 - \left(\frac{n}{k_0 r}\right)^2} - R \sqrt{1 - \left(\frac{n}{k_0 R}\right)^2} + \frac{n}{k_0} \left(\vartheta + \arcsin \frac{n}{k_0 r} - \arcsin \frac{n}{k_0 R}\right) + \alpha. \quad (94)$$

Uwzględniając dotychczasowe oznaczenia, widzimy, że zerowe przybliżenie φ_0 ma postać

$$\varphi_0 = AH_n^{(1)}(k_0 R) \exp \left\{ i n \left(\vartheta + \arcsin \frac{n}{k_0 r} - \arcsin \frac{n}{k_0 R} \right) + i k_0 \left[r \sqrt{1 - \left(\frac{n}{k_0 r}\right)^2} - R \sqrt{1 - \left(\frac{n}{k_0 R}\right)^2} \right] \right\}, \quad (95)$$

Związek między zmianami fazy w tym przybliżeniu i w rozwiązaniu ścisłym (79) jest widoczny z asymptotycznego wzoru dla funkcji Hankela

$$H_n^{(1)} \approx \sqrt{\frac{2}{\pi \varrho}} \exp \left[i \left(\varrho - \frac{2n+1}{4} \pi \right) \right]. \quad (96)$$

Niestety, jak widać, radialne zmiany amplitudy w zerowym przybliżeniu nie zostały uchwycone. Obliczamy przybliżenie pierwsze. W tym celu obliczamy najpierw $\nabla^2 {}_0A$. Po elementarnych rachunkach otrzymujemy

$$\nabla^2 {}_0A = \frac{1}{\sqrt{r^2 - \left(\frac{n}{k_0}\right)^2}}. \quad (97)$$

Ponieważ wyrażenie to zależy tylko od r , zatem, aby otrzymać funkcję p (według oznaczenia 64 poprzedniego rozdziału), należy wyrażenie to pomnożyć przez taki stały czynnik, aby funkcja p na krzywej Σ przyjmowała wartość 1. Czynnik ten oznaczamy przez β^2

$$\beta^2 = \sqrt{R^2 - \left(\frac{n}{k_0}\right)^2}. \quad (98)$$

Otrzymujemy teraz pierwsze przybliżenie

$$\varphi_1 = \frac{\beta}{\sqrt[4]{r^2 - \left(\frac{n}{k_0}\right)^2}} \varphi_0. \quad (99)$$

Biorąc pod uwagę wyrażenie (96) widzimy, że pierwsze przybliżenie oddaje asymptotyczny charakter zmian amplitudy.

Należy tu jednak zastrzec, że R powinno być dostatecznie duże, a mianowicie powinna być spełniona nierówność

$$R > \frac{n}{k_0}, \quad (100)$$

gdyż w przeciwnym przypadku dla pewnego r mianownik w wyrażeniu po prawej stronie równania (99) stawałby się równy zeru. Można wykazać,

że im większe jest R w stosunku do $\frac{n}{k_0}$, tym lepiej φ jest aproksymowane przez φ_1 . Sytuacja taka wydaje się dość naturalna, gdyż przybliżenia o charakterze optycznym powinny odznaczać się tym większą dokładnością, im mniej gwałtowne są zmiany pola narzucone przez warunki początkowe.

Interesujący jest fakt, że warunek brzegowy (3), który potem grał rolę warunku początkowego, nie określa jednoznacznie rozwiązania równania Helmholtza (potrzebny jest jeszcze warunek regularności w nieskończoności i warunek wypromieniowania), a zatem wskazane postępowanie aproksymacyjne wybrało samo pożądane rozwiązanie.

Warunek wypromieniowania, jak wiadomo, ma na celu wybranie rozwiązania odpowiadającego polu wychodzącemu z obszaru źródeł i rozchodzącemu się do nieskończoności, a nie na odwrót.

Problem Cauchy dla równania (nieliniowego) $(\nabla_0 A)^2 = 1$ jest niejednoznaczny, dopuszcza istnienie powierzchni ekwifazowych ${}_0A = \text{const}$ o fazie malejącej w miarę oddalania się do nieskończoności i o fazie rosnącej. Odpowiada temu wybór znaku pierwiastka w równaniach (40) i (41).

WYKAZ LITERATURY

1. Bohm D.: Phys. Rev. 85, 166, 180. 1952.
2. Landau-Lifszyc: *Kwantowa mechanika*. Gosizdat, Moskwa, 1948, s. 187.

K. БОХЕНЭК, И. ПЛЕБАНСКИ

О МЕТОДАХ МИКРОВОЛНОВОЙ ОПТИКИ

Резюме

Предложены два метода аппроксимации электромагнитных полей, являющиеся развитием методов геометрической оптики. Целью этих методов является облегчение анализа проблем связанных с микроволновой техникой. Представленные в настоящей работе два подхода к вопросу отличаются от тех, которые авторам известны из доступной им литературы. Не предпреляя правильности этих подходов, авторы полагают, что уместна разработка разных методов, которых применение будет сообразно приносимой ими пользе.

Первый метод отыскивает решение уравнения $(\Delta + k^2)u = 0$ в виде $u = Ae^{ik_0 L}$.

Выражение $\sqrt{1 + \frac{1}{k_0^2} \frac{\Delta A}{A}}$ играет роль действительного коэффициента преломления и может в случае надобности быть определено с точностью требуемого порядка относительно $\frac{1}{k_0}$. Эта интерпретация аналогична интерпретации квантовой механики Д. Боме. Краевая задача для уравнения $(\nabla + k^2)u = 0$ преобразована на задачи Коши для ряда уравнений выражающих последовательные члены рядов A и L относительно k_0^{-1} .

Второй метод за исходную точку принимает метод В. К. Б. т. е. отыскание решения уравнения $(\Delta + k^2)u = 0$ в виде $u = e^{ik_0 A}$, где функция A комплексная и зависит от k_0 как

$$A = \sum_{n=0}^{\infty} \left(\frac{1}{ik_0} \right)^n {}_n A.$$

Это ведёт к некоторым рядам задач Коши для функции ${}_n A$.

Первое приближение (со знаком 0) так относится к неоднородным плоским волнам, как геометрическая оптика к обыкновенным плоским волнам. Обсуждаемые проблемы Коши для комплексных функций дают значительную эффективность вычислений.

Оба метода пояснены примерами расчётов.

K. BOCHENEK, J. PLEBAŃSKI

ON METHODS OF MICROWAVE OPTICS

Summary

In this paper, two approximation methods developed on the basis of geometrical optics are proposed for treating electromagnetic fields. The purpose of these methods is to facilitate the analysis of microwave problems. The two processes presented here are different from those found in the literature available to the authors.

Apart from the question of the exclusive correctness of these particular concepts, the authors believe that it is advisable, to work out various methods whose scope of application will be then established by the benefit which their use may bring in practice.

According to the first method, a solution of the equation $(\Delta + k^2)u = 0$ is sought in the form: $u = Ae^{ik_0 L}$.

The expression

$$V = \sqrt{1 + \frac{1}{k_0^2} \frac{\Delta A}{A}}$$

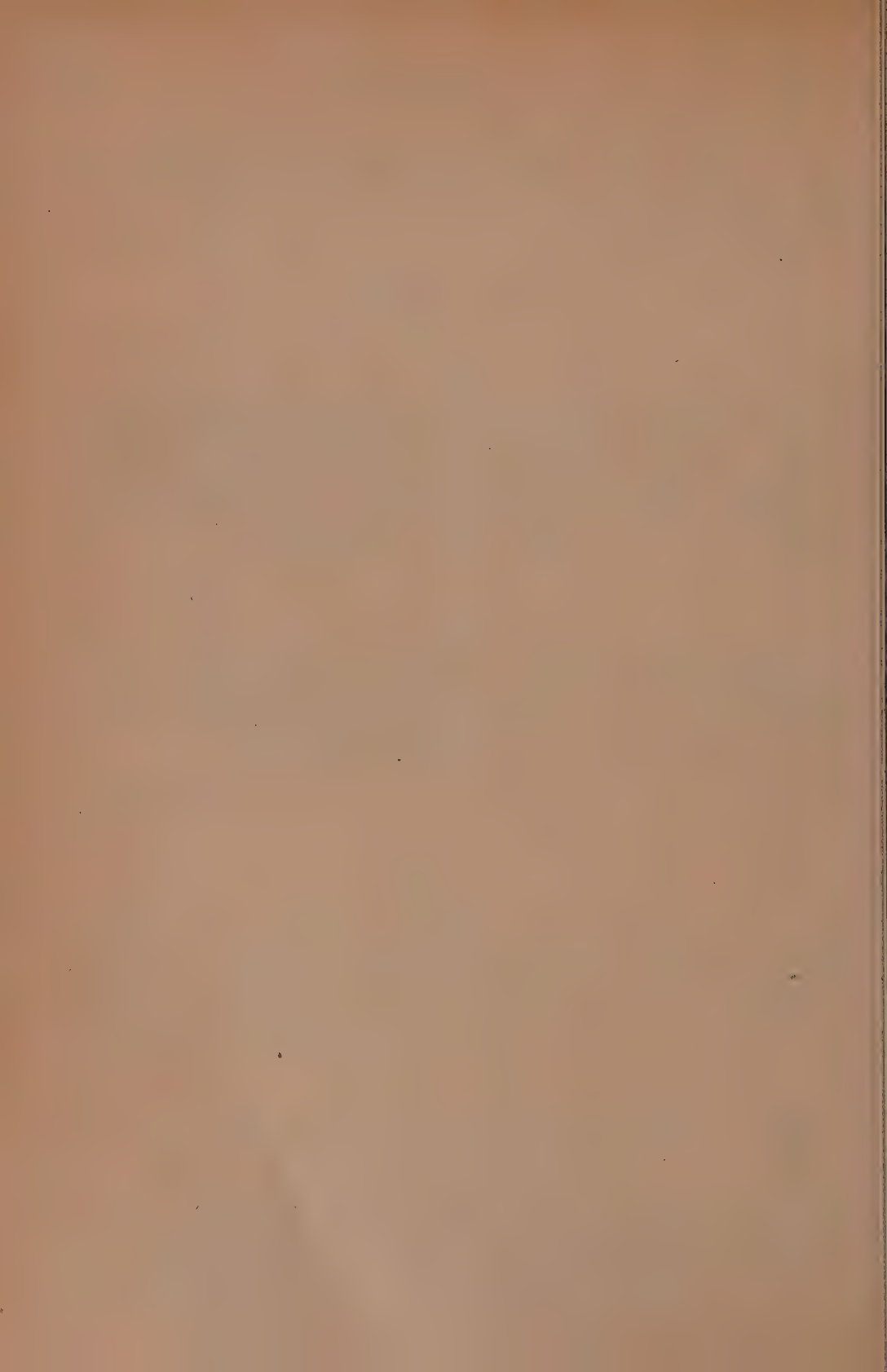
stands for an effective coefficient of refraction, and, if required can be determined to an accuracy of a small order of magnitude with respect to $\frac{1}{k_0}$. This interpretation is analogous to the quantum mechanical interpretation of D. Bohm. The boundary value problem for the equation $(\Delta + k^2)u = 0$ is transformed into a Cauchy problem for the set of equations for the successive term of the expansion of A and L in k_0^{-1} .

The second method takes the W. K. B. approximation as a starting point. This means that a solution of the equation $(\Delta + k^2)u = 0$ is sought in the form $u = e^{ik_0 A}$ where the function A is a complex one and depends on k_0 as

$$A = \sum_{n=0}^{\infty} \left(\frac{1}{ik_0} \right)^n {}_n A.$$

This leads to the set of Cauchy's problems for the function ${}_n A$. The first approximation (with the index 0) has the same relation to non-homogeneous plane waves as geometrical optics has to ordinary plane waves. The above-mentioned Cauchy problems for complex functions can be in general effectively solved.

Both methods are illustrated by sample calculations.



621.396 6772

S. MANCZARSKI

Redukcja listków bocznych w antenach kierunkowych

*Rękopis dostarczono 13. 12. 1955 **

W pracy omówiono zagadnienie redukcji listków bocznych w charakterystykach promieniowania trzech zasadniczych typów radiokomunikacyjnych anten krótkofalowych: anten rombów, ścianowych synfazowych oraz wzdłużnych. Na tle obcych opracowań metod redukcji listków bocznych autor przedstawia własną metodę, opartą na zastosowaniu anteny wzdłużnej z tzw. falą stojąco-bieżącą. Głównymi zaletami nowej anteny jest jej prostota, wysoka sprawność oraz możliwość obciążania jej dużą mocą. Anteny z falą stojąco-bieżącą są stosowane w Polsce w radiofonicznej służbie krótkofalowej.

W ciągu ostatnich lat propagowana jest przez Międzynarodowy Komitet Doradczy CCIR akcja redukcji listków bocznych w charakterystykach promieniowania krótkofalowych anten radiokomunikacyjnych. Akcja ta odpowiada ogólnej tendencji złagodzenia wzajemnych zakłóceń radiostacji przez usunięcie lub zmniejszenie promieniowania energii w kierunkach niepożądanych. W związku z tym wymieniona akcja dotyczy przede wszystkim anten nadawczych.

Tak zwane listki boczne występują w silniejszym lub słabszym stopniu przy wszelkich rodzajach anten o zwiększonej kierunkowości i redukcja ich jest, praktycznie biorąc, zagadnieniem niełatwym. W radiokomunikacji krótkofalowej zagadnienie to jest najbardziej aktualne w odniesieniu do trzech następujących typów anten kierunkowych:

- a. anten rombów (rhombic antennas),
- b. anten ścianowych synfazowych (broadside arrays),
- c. anten wzdłużnych (endfire arrays).

Ze względu na swą szerokopasmowość najczęściej stosowane są obecnie, zwłaszcza w celach nadawczych, anteny rombowe. Niestety, anteny te w porównaniu z innymi typami anten kierunkowych odznaczają się stosunkowo znacznymi rozmiarami listków bocznych. Systemem zmniejszającym do pewnego stopnia listki boczne jest radziecki układ podwójnego

* Praca referowana przez autora na zjeździe naukowym w Ciechocinku we wrześniu 1955 r.

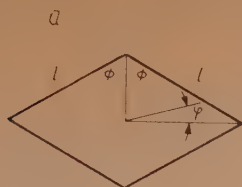
rombu Ajzenberga [1]. Wyniki, uzyskiwane za pomocą tego rombu, pokazano na rys. 1, 2 i 3, gdzie zestawiono charakterystyki promieniowania rombu pojedynczego a. i podwójnego b. Charakterystyki te podano w układzie współrzędnych prostokątnych. Jak widać z tych rysunków, poważniejszą redukcję listków bocznych osiąga się jedynie w pobliżu pewnej określonej częstotliwości; na przykład dla rombu podwójnego o boku $l=4\lambda_0$, gdzie λ_0 jest optymalną długością fali, minimalny poziom listków bocznych przypada dla $\lambda \approx 1,33 \lambda_0$ (rys. 2).

Należy zauważyć, że przedstawione na trzech wymienionych rysunkach charakterystyki promieniowania rombów obliczone są na podstawie wzorów podanych na rys. 1 i odnoszących się, ściśle biorąc, do promieniowania w wolnej przestrzeni. Wykresy te mogą jednak z dostatecznym przybliżeniem ilustrować również charakterystyki promieniowania dla niewielkich kątów elewacji w przypadku rombów umieszczonych nad ziemią. Przybliżenie takie stosuje się dość często w obliczeniach antenowych dla uproszczenia rachunków.

W odniesieniu do synfazowych anten ścianowych znanym sposobem redukcji listków bocznych jest wytworzenie nierównomiernego rozkładu prądu w ścianie. Rozkład ten dobiera się w taki sposób, aby maksymalny prąd płynął w środkowych elementach ściany. Najprostszy przykład zastosowania powyższej metody pokazano na rys. 4 dla anteny czteroelementowej z reflektorem biernym w postaci ekranu. Na rysunku tym zestawiono charakterystyki poziome promieniowania anteny w wolnej przestrzeni z równomiernym rozkładem prądu a. i nierównomiernym — b., przy czym w drugim przypadku prąd w środkowych elementach przyjęto o 80% większy od prądu w elementach skrajnych. Pod wykresem podano wzory ułatwiające obliczenie odnośnych krzywych. Wartości liczbowe współczynników wchodzących w skład wymienionych wzorów odczytuje się z tablicy na rys. 5, która została zestawiona przez autora [6] dla szybkiego obliczania charakterystyk promieniowania zarówno anten synfazowych jak i wzdłużnych.

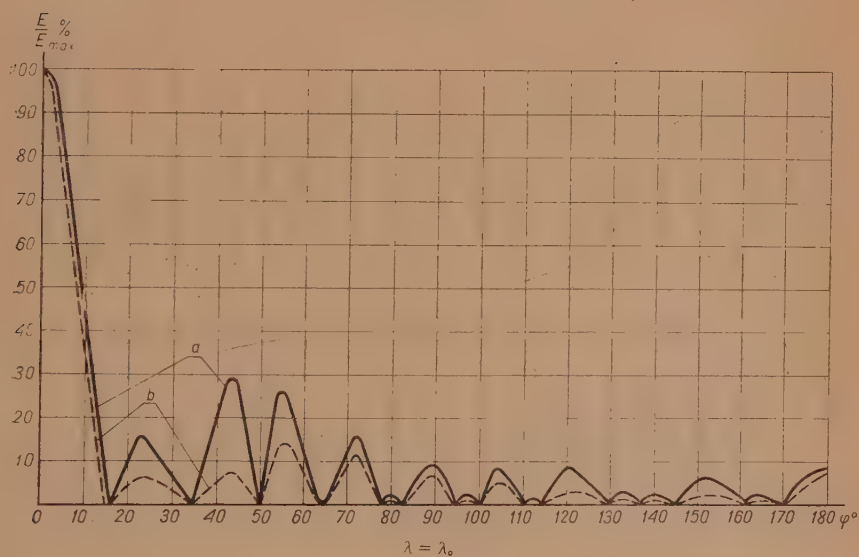
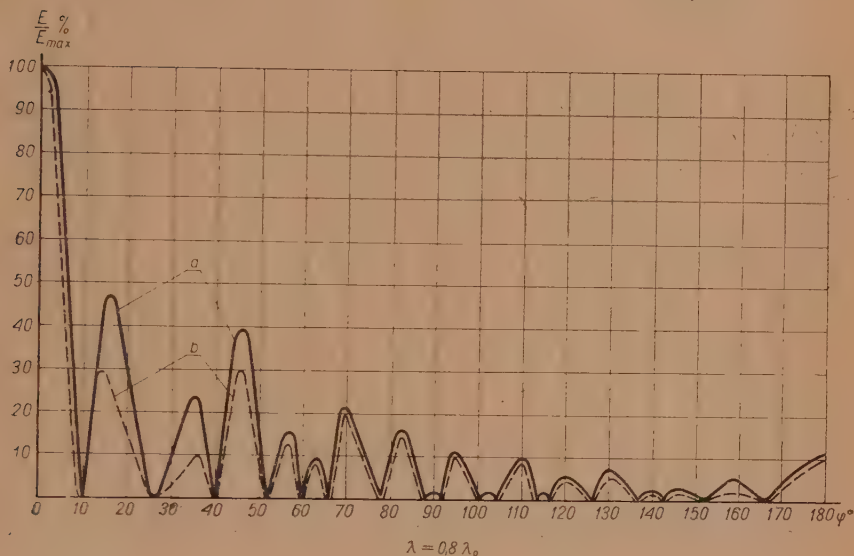
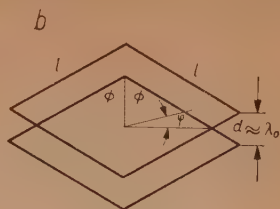
Jak widać z rys. 4, zmniejszenie listka bocznego (w danym przypadku występują tylko dwa listki boczne, po jednym z każdej strony listka głównego) powoduje równocześnie nieznaczne poszerzenie listka głównego.

Ogólnie biorąc, ilość listków bocznych jest tym większa, im większa liczba elementów wchodzi w skład danej anteny ścianowej. Przy występowaniu wielu listków bocznych po każdej stronie listka głównego powstaje zagadnienie, jak należy ustalić rozkład prądu w poszczególnych elementach, ażeby otrzymać równocześnie zredukowany poziom wszystkich listków bocznych. Pod względem teoretycznym zagadnienie to zostało rozwiązane w roku 1946 przez C. L. Dolpha [2], który wykazał, że przy zastosowaniu wielomianu Czebyszewa możliwe jest zaprojektowanie



$$\phi = 65^\circ$$

$$\lambda_0 = \frac{l}{4}$$

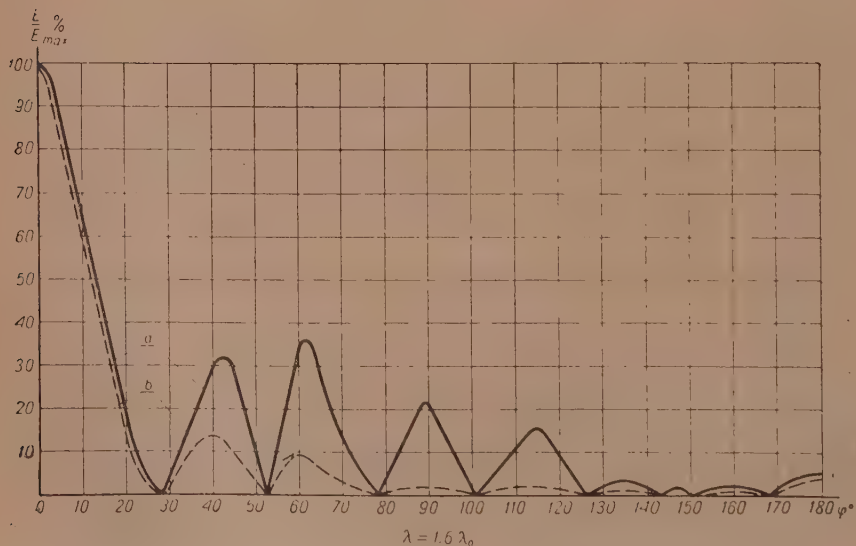
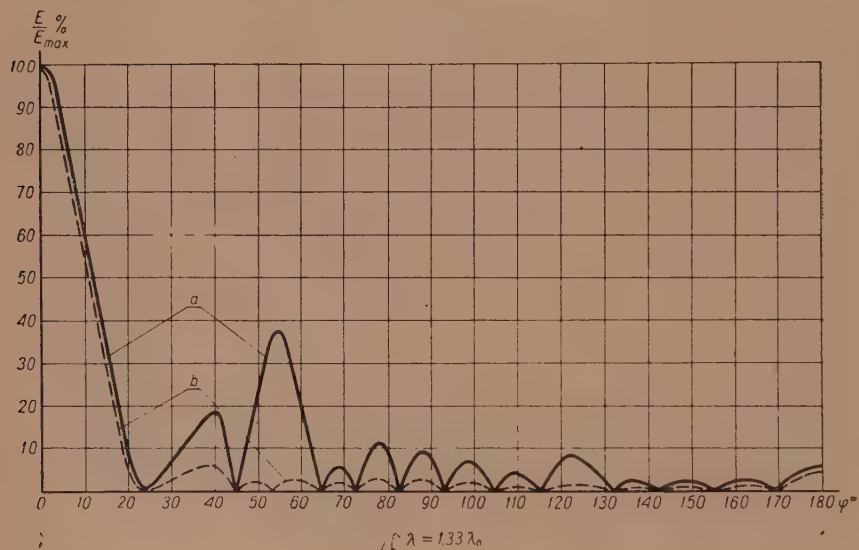
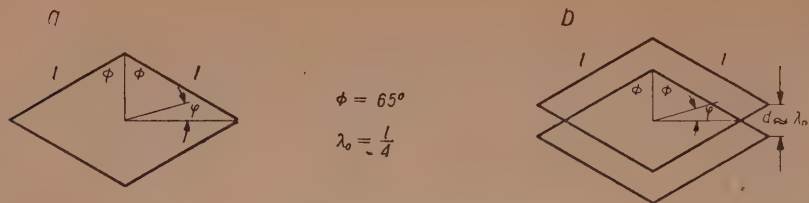


$$F_a(\varphi) = \left[\frac{\cos(\phi + \varphi)}{1 - \sin(\phi + \varphi)} + \frac{\cos(\phi - \varphi)}{1 - \sin(\phi - \varphi)} \right] \cdot \sin \left\{ \pi \frac{l}{\lambda} [1 - \sin(\phi + \varphi)] \right\} \sin \left\{ \pi \frac{l}{\lambda} [1 - \sin(\phi - \varphi)] \right\}$$

$$F_b(\varphi) = F_a(\varphi) \cos \left(\pi \frac{d}{\lambda} \sin \varphi \right)$$

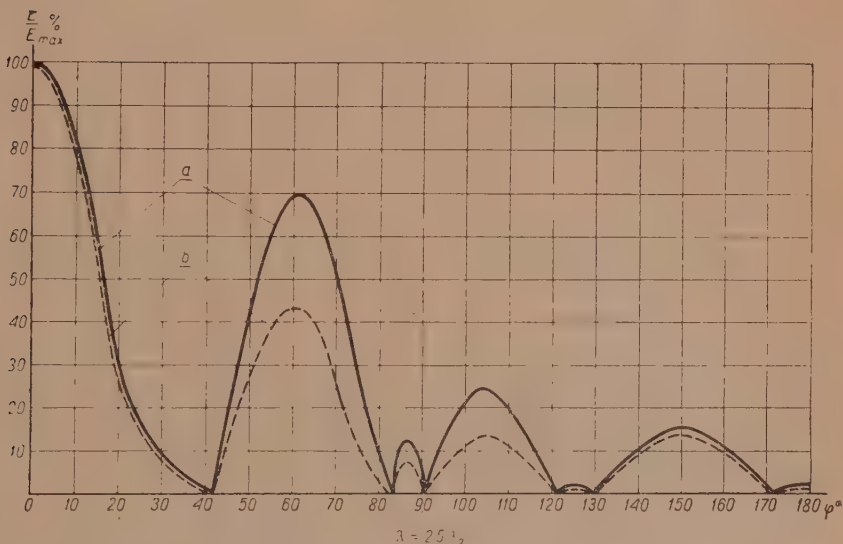
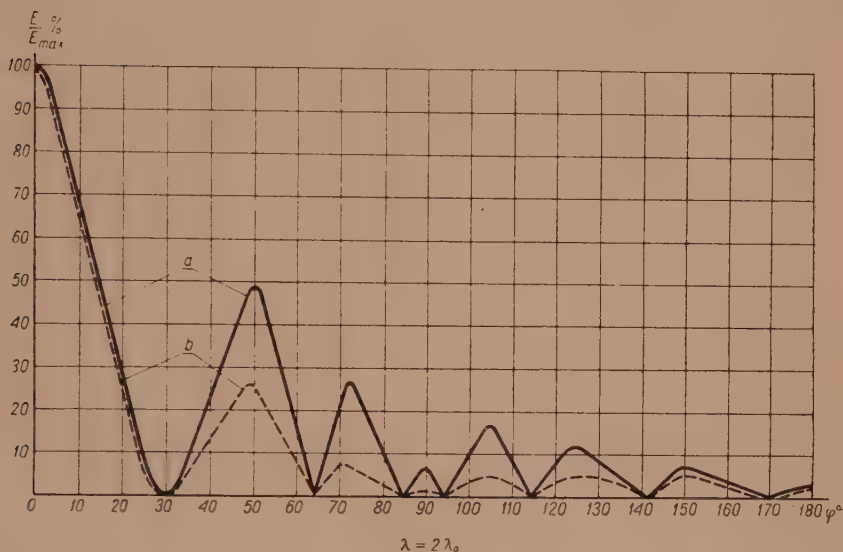
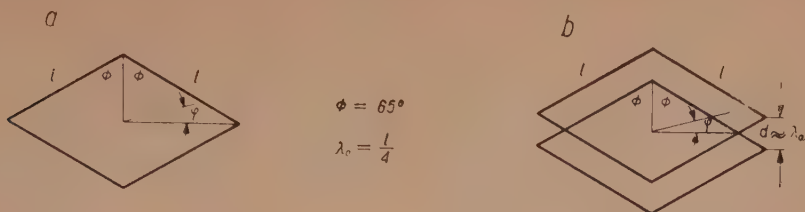
Rys. 1. Charakterystyki poziome anten rombów pojedynczej a. i podwójnej b.

Wg G. Z. Ajzenberga



Rys 2. Charakterystyki poziome anten rombów pojedynczej a. i podwójnej b.

Wg G. Z. Ajzenberga



Rys. 3. Charakterystyki poziome anten rombów pojedynczej *a*. i podwójnej *b*.

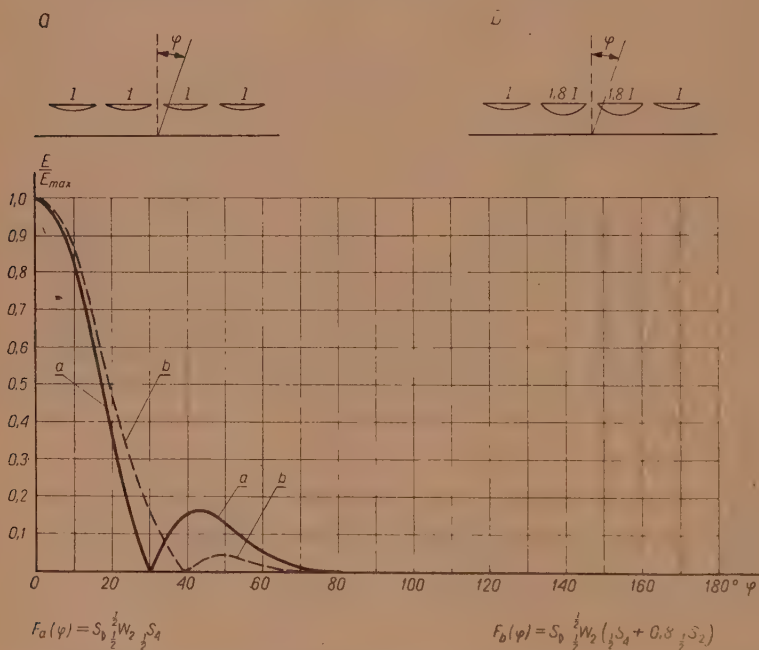
Wg G. Z. Ajzenberga

takiego rozkładu prądu w antenie ścianowej, przy którym poziom wszystkich listków bocznych stanowi pewną stałą, z góry obraną wartość. Wybrany poziom listków bocznych, przy danych rozmiarach anteny, określa szerokość listka głównego, która zwiększa się nieco w miarę zmniejszania poziomu listków bocznych.

Jak wiadomo, wielomian Czebyszewa rzędu N zdefiniowany jest przez następujące wyrażenia:

$$T_N(z) = \cos(N \arccos z), \quad \text{gdzie } |z| \leq 1, \quad (1)$$

$$T_N(z) = \cosh(N \operatorname{arccosh} z), \quad \text{gdzie } |z| > 1. \quad (2)$$



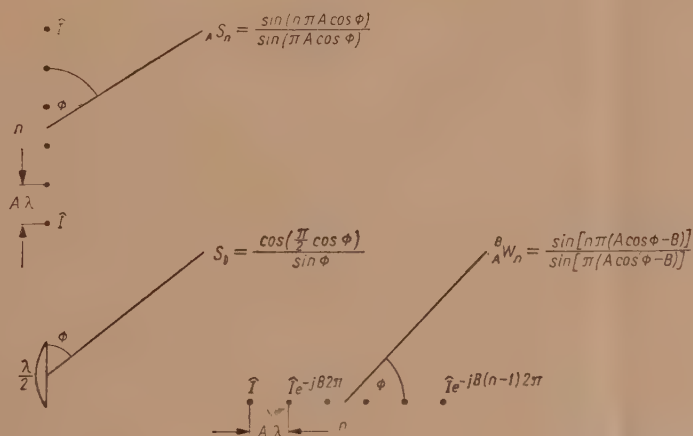
Rys. 4. Charakterystyki poziome synfazowych anten ścianowych z równomiernym a. i nierównomiernym b. rozkładem prądu.

Wg S. Manczarskiego

Dalsze przekształcenie tych wyrażeń w zastosowaniu do charakterystyk promieniowania anten można znaleźć w pracach [2] lub [3]. Redukcja listków bocznych pod względem matematycznym rozwiązuje się w myśl tych prac w sposób całkowicie zadowalający.

W praktyce jednak realizacja metody Dolpha natrafia na bardzo poważne trudności, zwłaszcza przy antenach ścianowych o znacznej liczbie elementów. Sprawa regulowania, bez strat energii, wartości prądu w poszczególnych elementach dużej anteny ścianowej nie przedstawia się prosto. Z tego względu w krótkofalowych antenach radiokomunikacyj-

S_D	$\frac{1}{4} S_2$	$\frac{1}{4} S_4$	$\frac{1}{4} S_8$	$1 S_2$	$\frac{3}{2} S_2$	$\frac{1}{2} S_2$	$\frac{1}{2} S_3$	$\frac{1}{2} S_4$	$\frac{1}{2} S_5$	$\frac{1}{2} S_6$	$\frac{1}{2} S_8$	Φ	$\frac{1}{4} W_2$	$\frac{1}{2} W_2$
0	1,41	0	0	-2	0	0	-1	0	1	0	0	0	2	2
0,080	1,42	0,020	-0,322	-2,00	-0,045	0,014	-1,00	-0,023	1,00	0,052	-0,052	5	2,00	2,00
0,141	1,43	0,068	-0,137	-2,00	-0,140	0,047	-0,998	-0,096	0,993	0,139	-0,208	10	2,00	2,00
0,205	1,45	0,154	-0,302	-1,99	-0,312	0,106	-0,989	-0,208	0,966	0,310	-0,407	15	2,00	2,00
0,275	1,48	0,311	-0,560	-1,96	-0,568	0,209	-0,957	-0,377	0,894	0,592	-0,698	20	2,00	1,99
0,346	1,52	0,442	-0,845	-1,91	-0,855	0,292	-0,916	-0,558	0,737	0,775	-0,931	25	2,00	1,98
0,418	1,53	0,635	-1,16	-1,83	-1,18	0,416	-0,827	-0,759	0,508	0,975	-1,008	30	1,99	1,96
0,490	1,60	0,903	-1,52	-1,69	-1,51	0,563	-0,684	-0,948	0,154	1,032	-0,798	35	1,98	1,92
0,557	1,65	1,18	-1,56	-1,48	-1,78	0,717	-0,486	-1,067	-0,283	0,855	-0,331	40	1,97	1,87
0,627	1,70	1,51	-1,83	-1,21	-1,96	0,888	-0,213	-1,076	-0,744	0,418	0,575	45	1,95	1,79
0,696	1,75	1,87	-1,62	-0,868	-1,99	1,06	0,128	-0,922	-1,03	-0,246	1,152	50	1,92	1,69
0,758	1,80	2,24	-1,03	-0,458	-1,81	1,24	0,539	-0,570	-1,25	-0,978	1,02	55	1,89	1,57
0,817	1,85	2,61	0	0	-1,41	1,41	1	0	-1	-1,414	0	60	1,85	1,41
0,868	1,89	2,98	1,43	0,480	-0,813	1,58	1,48	0,757	-0,287	-1,214	-1,34	65	1,80	1,232
0,914	1,93	3,30	3,15	0,952	0,084	1,72	1,95	1,64	0,859	-0,163	-1,79	70	1,74	1,024
0,955	1,96	3,62	4,96	1,37	0,684	1,83	2,37	2,53	2,36	1,63	-0,282	75	1,67	0,790
0,978	1,98	3,82	6,54	1,71	1,37	1,93	2,71	3,29	3,63	3,71	3,04	80	1,59	0,538
0,995	1,99	3,98	7,66	1,93	1,86	1,98	2,93	3,82	4,65	5,37	6,50	85	1,51	0,272
1	2	4	8	2	2	2	3	4	5	6	8	90	1,41	0
												95	1,31	-0,272
												100	1,21	-0,538
												105	1,10	-0,790
												110	0,988	-1,024
												115	0,876	-1,232
												120	0,765	-1,41
												125	0,657	-1,57
												130	0,550	-1,69
												135	0,456	-1,79
												140	0,364	-1,87
												145	0,283	-1,92
												150	0,209	-1,96
												155	0,147	-1,98
												160	0,091	-1,99
												165	0,054	-2,00
												170	0,021	-2,00
												175	0,006	-2,00
												180	0	-2

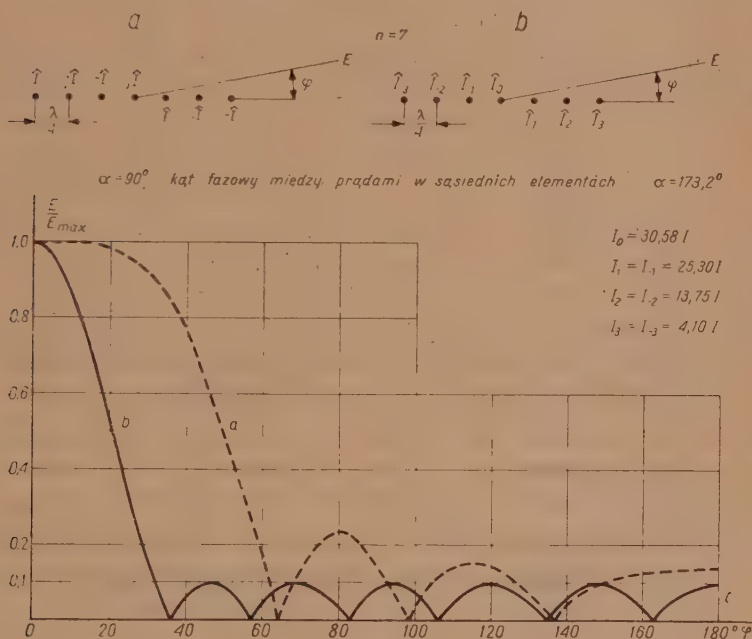


Rys. 5. Współczynniki dla obliczania charakterystyk promieniowania anten synfazowych i wzdlużnych.

W_2	$\frac{1}{3}W_2$	$\frac{1}{2}W_2$	$\frac{2}{3}W_2$	$\frac{5}{6}W_2$	$\frac{1}{6}W_2$	$\frac{1}{3}W_2$	$\frac{1}{3}W_2$	$\frac{2}{3}W_2$	$\frac{5}{6}W_2$	$\frac{1}{2}W_2$	$\frac{1}{4}W_4$	$\frac{1}{4}W_8$	$\frac{1}{5}W_8$
73	-1	0	1	1,73	-1,00	-1,73	-2	-1,73	-1,00	0	4	8	6,08
72	-0,976	0,024	1,02	1,74	-1,04	-1,75	-2,00	-1,71	-0,960	-0,048	4,00	8,00	6,02
68	-0,916	0,094	1,11	1,78	-1,12	-1,80	-1,99	-1,66	-0,877	-0,191	4,00	8,00	5,86
62	-0,810	0,214	1,18	1,83	-1,26	-1,87	-1,97	-1,55	-0,717	-0,425	3,99	7,95	5,575
51	-0,656	0,376	1,31	1,89	-1,45	-1,95	-1,92	-1,38	-0,467	-0,740	3,97	7,8	5,162
37	-0,454	0,580	1,46	1,95	-1,64	-1,99	-1,81	-1,14	-0,164	-1,111	3,948	7,56	4,61
17	-0,206	0,816	1,62	1,99	-1,83	-1,99	-1,62	-0,813	0,209	-1,492	3,89	7,1	3,92
922	0,088	1,08	1,77	2	-1,96	-1,89	-1,31	-0,382	0,651	-1,814	3,8	6,41	3,095
614	0,420	1,34	1,90	1,96	-2	-1,68	-0,908	0,105	1,09	-1,990	3,672	5,444	2,172
254	0,772	1,59	1,98	1,84	-1,89	-1,31	-0,382	0,651	1,51	-1,928	3,49	4,225	1,198
150	1,13	1,80	1,99	1,65	-1,61	-0,801	0,223	1,19	1,83	-1,563	3,253	2,82	0,24
576	1,46	1,95	1,92	1,37	-1,41	-0,167	0,852	1,64	1,99	-0,892	2,952	1,356	-0,62
00	1,73	2	1,73	1,00	-0,518	0,518	1,41	1,93	1,93	0	2,612	0	-1,293
39	1,92	1,94	1,44	0,554	0,209	1,18	1,83	1,99	1,62	0,935	2,222	-1,067	-1,71
70	2	1,76	1,05	0,054	0,927	1,69	2	1,77	1,07	1,675	1,78	-1,694	-1,83
92	1,95	1,45	0,570	-0,464	1,53	1,97	1,88	1,29	0,347	1,997	1,32	-1,814	-1,652
	1,75	1,04	0,042	-0,962	1,92	1,95	1,46	0,578	-0,457	1,774	0,859	-1,468	-1,223
94	1,43	0,542	-0,494	-1,40	1,98	1,56	0,743	-0,285	-1,23	1,041	0,411	-0,793	-0,635
73	1,00	0	-1,00	-1,73	1,73	1,00	0	-1,00	-1,73	0	0	0	0
40	0,494	-0,542	-1,43	-1,94	1,23	0,285	-0,743	-1,56	-1,98	-1,041	-0,358	0,691	0,57
962	-0,042	-1,04	-1,75	-2	0,457	-0,578	-1,46	-1,95	-1,92	-1,774	-0,651	1,111	0,982
464	-0,570	-1,45	-1,95	-1,92	-0,347	-1,29	-1,88	-1,97	-1,53	-1,997	-0,87	1,195	1,188
054	-1,05	-1,76	-2	-1,70	-1,07	-1,77	-2	-1,69	-0,927	-1,675	-1,011	0,964	1,174
554	-1,44	-1,94	-1,92	-1,39	-1,62	-1,99	-1,83	-1,18	-0,209	-0,935	-1,08	0,52	0,981
00	-1,73	-2	-1,73	-1,00	-1,93	-1,93	-1,41	-0,518	0,518	0	-1,082	0	0,66
37	-1,92	-1,95	-1,46	-0,576	-1,99	-1,64	-0,852	0,167	1,14	0,892	-1,03	-0,472	0,28
65	-1,99	-1,80	-1,13	-0,150	-1,83	-1,19	-0,223	0,801	1,61	1,563	-0,938	-0,814	-0,096
84	-1,98	-1,59	-0,772	0,254	-1,51	-0,651	0,382	1,31	1,89	1,928	-0,818	-0,99	-0,425
96	-1,90	-1,34	-0,420	0,614	-1,09	-0,105	0,908	1,68	2	1,990	-0,683	-1,012	-0,683
	-1,77	-1,08	-0,088	0,922	-0,651	0,382	1,31	1,89	1,96	1,814	-0,543	-0,916	-0,859
99	-1,62	-0,816	0,206	1,17	-0,209	0,813	1,62	1,99	1,83	1,492	-0,411	-0,751	-0,964
95	-1,46	-0,580	0,454	1,37	0,164	1,14	1,81	1,99	1,64	1,111	-0,292	-0,557	-1,01
89	-1,31	-0,376	0,656	1,51	0,467	1,38	1,92	1,95	1,45	0,740	-0,188	-0,370	-1,02
83	-1,18	-0,214	0,810	1,62	0,717	1,55	1,98	1,87	1,26	0,425	-0,107	-0,212	-1,005
78	-1,11	-0,094	0,916	1,68	0,877	1,66	1,99	1,80	1,12	0,191	-0,048	-0,095	-0,986
74	-1,024	-0,02	0,976	1,72	0,960	1,71	2,00	1,75	1,04	0,048	-0,012	-0,024	-0,97
73	-1,00	0	1,00	1,73	1,00	1,73	2	1,75	1,00	0	0	0	-0,963

nych (to znaczy, nie biorąc pod uwagę radaru) stosuje się jedynie bardzo dalekie przybliżenie do rozkładu, odpowiadającego metodzie Dolpha. Ma to miejsce w antenach ścianowych zasilanych centralnie w jednym punkcie, na przykład w antenie Chireixa lub antenach budowanych przez polską administrację pocztową.

W praktyce nie lepiej przedstawia się poruszona sprawa w odniesieniu do anten wzdłużnych. I tutaj, podobnie jak w antenach ścianowych, istnieje teoretyczna możliwość dowolnego zmniejszenia listków bocznych



Rys. 6. Podstawowe charakterystyki promieniowania anteny wzdłużnej normalnej *a*. i ze zredukowanymi listkami bocznymi *b*. przy zastosowaniu wielomianu Czebyszewa.

Wg R. H. Duhamela

przez wytworzenie nierównomiernego rozkładu prądu w antenie wzdłużnej. W roku 1953 R. H. Duhamel [3] wykazał, że przez zastosowanie wielomianu Czebyszewa można doprowadzić w antenie wzdłużnej wszystkie listki boczne do z góry wybranego, stałego poziomu. Ilustrację tego rozwiązania przedstawiono na rys. 6, gdzie dla przykładu rozważona jest antena wzdłużna siedmioelementowa (nieparzysta liczba elementów daje pewne ułatwienia matematyczne przy stosowaniu wielomianu Czebyszewa). Na rysunku tym zestawiono podstawowe charakterystyki promieniowania anteny wzdłużnej normalnej *a*. i ze zredukowanymi listkami bocznymi *b*. przy zastosowaniu wielomianu Czebyszewa.

Termin *podstawowe charakterystyki promieniowania* oznacza, że charakterystyki te zdjęto w wolnej przestrzeni w płaszczyźnie prostopadłej do elementów, z których składa się dana antena wzdłużna. W obu rozważanych przypadkach a. i b. odległość między sąsiednimi elementami jest ta sama i wynosi $\frac{\lambda}{4}$. Jednakże w przypadku a. prądy we wszystkich elementach posiadają tę samą wartość I , a kąt fazowy między prądami w sąsiednich elementach wynosi $\alpha = 90^\circ$ odpowiednio do odległości $\frac{\lambda}{4}$ (stąd nazwa *antena wzdłużna normalna*). W przypadku b. prądy w poszczególnych elementach są różne pod względem wartości i wykazują minimum w elementach skrajnych, a maksimum w elemencie środkowym i kąt fazowy między prądami w sąsiednich elementach wynosi $\alpha = 173,2^\circ$.

Rozkład prądu i kąt fazowy α wybrano w przypadku b. przy zastosowaniu wielomianu Czebyszewa w taki sposób, aby wszystkie listki boczne osiągnęły ten sam poziom, wynoszący 10% natężenia pola wytwarzanego przez listek główny. Na rys. 6 podano prądy I_0, I_1, I_2, I_3 dla przypadku b., obliczone w funkcji prądu I , wziętego z przypadku a. w taki sposób, aby natężenie pola listka głównego dawało w obu przypadkach tę samą wartość. Z przedstawionego na tym rysunku liczbowego zestawienia prądów, wynikają dwa bardzo ważne wnioski:

1. oporność promieniowania anteny w przypadku b. musi być znacznie mniejsza niż w przypadku a.,

2. prądy w przypadku b. muszą być doregulowane, jak podaje Duhamel [3] z dużą precyzją, rzędu 0,1% w stosunku do prądu maksymalnego.

Pierwszy z tych wniosków związany jest przede wszystkim z widocznym na rys. 6 dość znacznym zwężeniem listka głównego w przypadku b., co zbliża ten przypadek do warunków panujących przy tak zwanych antenach superkierunkowych (z tym jednak zastrzeżeniem, że rozkład prądu w antenie superkierunkowej, obliczony na przykład metodą Schelkunoffa, różni się od rozkładu prądu obliczonego metodą Duhamela). Jest rzeczą jasną, że nadmierne zmniejszenie oporności promieniowania jest okolicznością niekorzystną dla anten nadawczych, zarówno ze względu na pogorszenie sprawności anteny, jak i z uwagi na przepięcia występujące przy dużych mocach. Równocześnie krzywa rezonansu anteny może okazać się wówczas zbyt ostra dla pewnych rodzajów transmisji.

Drugi z wymienionych wniosków stwarza w praktyce niezwykle duże trudności, zwłaszcza jeżeli precyzyjna regulacja amplitudy i fazy poszczególnych prądów ma być zrealizowana bez straty promieniowanej energii. W związku z tymi trudnościami piękna pod względem teoretycznym praca Duhamela nie znalazła dotąd w radiokomunikacji praktycznego zastosowania. To samo da się powiedzieć o pracach teoretycznych, dotyczących

anten wzdłużnych, Schelkunoffa [9] i Gowarda [4]. Pierwszy z nich stworzył metodę obliczania anten superkierunkowych ze zredukowanymi, ale niedoprowadzonymi do wspólnego poziomu listkami bocznymi, drugi osiągnął podobny rezultat przez zmodyfikowanie i uzupełnienie metody Hansena i Woodyarda [5] metodą Woodwarda [10].

W rozważaniach nad antenami wzdłużnymi nie wspomniano dotychczas o praktycznie używanym sposobie redukcji listków bocznych przez zastosowanie podwójnych anten dwustronnie grzebieniowych (array of fishbone antennas). Anteny tego typu nie wykazują omówionych niedogodności, ale ze względu na niską sprawność mogą być stosowane jedynie jako anteny odbiorcze. Redukcja listków bocznych opiera się tu na tej samej zasadzie jak w rombach podwójnych (patrz wzór $F_b(\varphi)$ na rys. 1); listki boczne otrzymuje się przy antenach grzebieniowych nieco mniejsze niż przy antenach rombowych.

W Polsce badania nad nadawczymi antenami wzdłużnymi o zredukowanych listkach bocznych rozpoczęto w roku 1950. Zwrócenie uwagi na antenę wzdłużną spowodowane było tą okolicznością, że anteny tego typu mają z punktu widzenia radiofonii krótkofalowej korzystniejszy kształt listka głównego niż anteny innych typów. Polskie Radio, które interesowało się tym zagadnieniem, postawiło dla anteny wzdłużnej następujące wymagania wstępne:

a. jak najlepsza koncentracja promieniowanej energii w żądanym kącie, czyli najodpowiedniejszy kształt listka głównego i redukcja listków bocznych,

b. możliwość obciążania anteny dużą mocą (rzędu kilkuset kilowatów),

c. wysoka sprawność anteny.

d. proste i tanie rozwiązanie zarówno schematowe jak i konstrukcyjne, nie wymagające skomplikowanych regulacji.

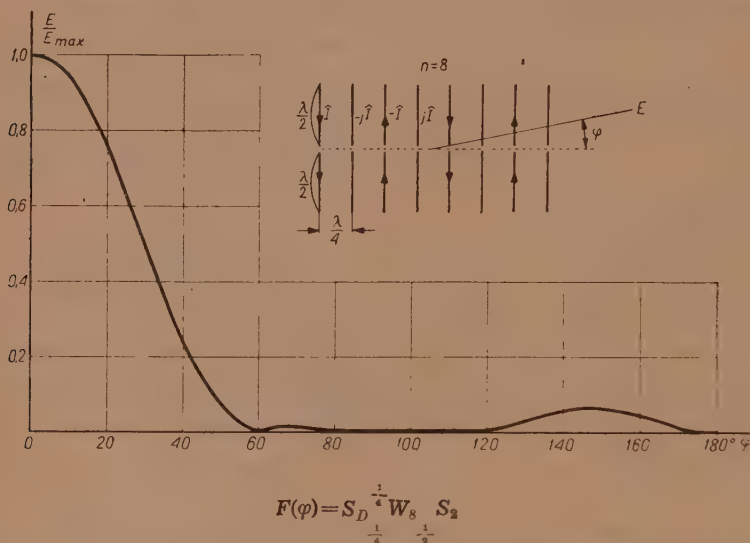
Wobec tych wymagań rozwiązanie koncepcyjne, jakie opracował autor [7], zostało oparte na antenie wzdłużnej normalnej, mającej dostatecznie wysoką oporność promieniowania, to jest na antenie z tak zwaną falą stojąco-bieżącą. Zastosowanie takiej fali stanowiło nowość w budowie wieloelementowych anten wzdłużnych, co pozwoliło na usunięcie pochłaniającego energię oporu końcowego, jaki używano zawsze w antenach z falą bieżącą; dzięki temu udało się otrzymać wysoką sprawność nowej anteny.

Antena wzdłużna z falą stojąco-bieżącą jest pomyślana w ten sposób, że w poszczególnych jej elementach występuje fala stojąca, ale pod względem promieniowania energii cała antena zachowuje się tak, jak gdyby istniała w niej fala bieżąca. Powyższy system anteny jest przeciwstawieniem innego systemu antenowego, również opracowanego przez autora [8]

pod nazwą anteny z falą bieżąco-stojącą, opartego na odmiennej zasadzie działania niż anteny wzdłużne.

Redukcja listków bocznych w rozpatrywanej antenie wzdłużnej z falą stojąco-bieżącą osiągana jest w znacznym stopniu dzięki kompensacyjnemu działaniu dwóch równoległych rzędów dipoli półfalowych, z których składa się omawiana antena.

Rozkład prądów i stosunki fazowe w antenie wzdłużnej z falą stojąco-



Rys. 7. Charakterystyka promieniowania w płaszczyźnie dipoli anteny wzdłużnej normalnej $j|8DD|1$ (w układzie współrzędnych prostokątnych).

Wg S. Manczarskiego

bieżącą wyjaśnia rys. 7. Rozważoną na tym rysunku antenę oznaczono symbolem $j|8DD|1$, gdzie

j — kąt fazowy 90° między prądami w sąsiednich elementach,

$8DD$ — 8 półfalowych dipoli podwójnych,

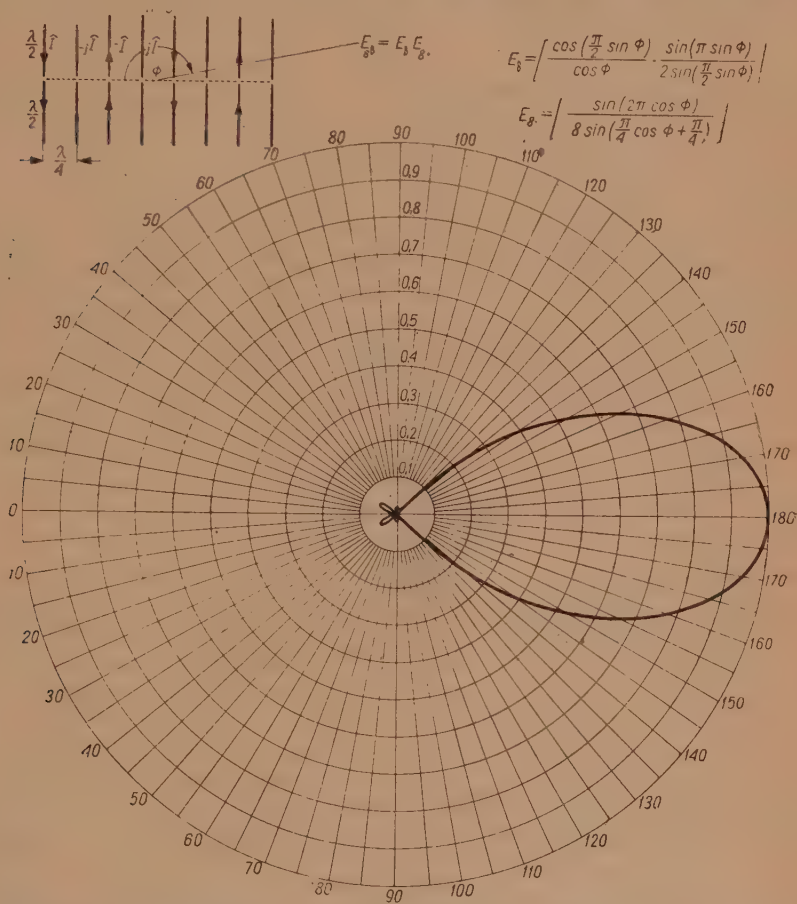
1 — współczynnik wskazujący, że antena jest normalna, to znaczy, że odległość między sąsiednimi elementami wynosi $\frac{\lambda}{4}$, co odpowiada kątowi fazowemu 90° wyrażonemu symbolem j .

Przedstawiona na rys. 7 charakterystyka promieniowania w płaszczyźnie dipoli anteny $j|8DD|1$ może być obliczona za pomocą wzoru podanego pod wykresem. Wchodzące w ten wzór współczynniki odczytuje się z tablicy na rys. 5. Jak widać z przebiegu rozważanej charakterystyki, listki boczne uległy tu dość znacznej redukcji, co jest widoczne szczególnie wyraźnie przez porównanie rys. 7 z rys. 6.

Dla lepszego jeszcze zobrazowania charakterystyki promieniowania

anteny wzdłużnej $j|8DD|1$ charakterystykę tę przedstawiono dodatkowo na rys. 8 w układzie współrzędnych biegunowych. Podane tam wzory na natężenie pola E w funkcji kąta Φ są tylko inną wersją wzorów figurujących na rys. 5, dostosowaną do odmiennego sposobu odmierzania kąta Φ .

Oporność promieniowania anteny $j|8DD|1$, odniesiona do prądu w pojedynczym dipolu półfalowym, wynosi w wolnej przestrzeni około 2450Ω ,

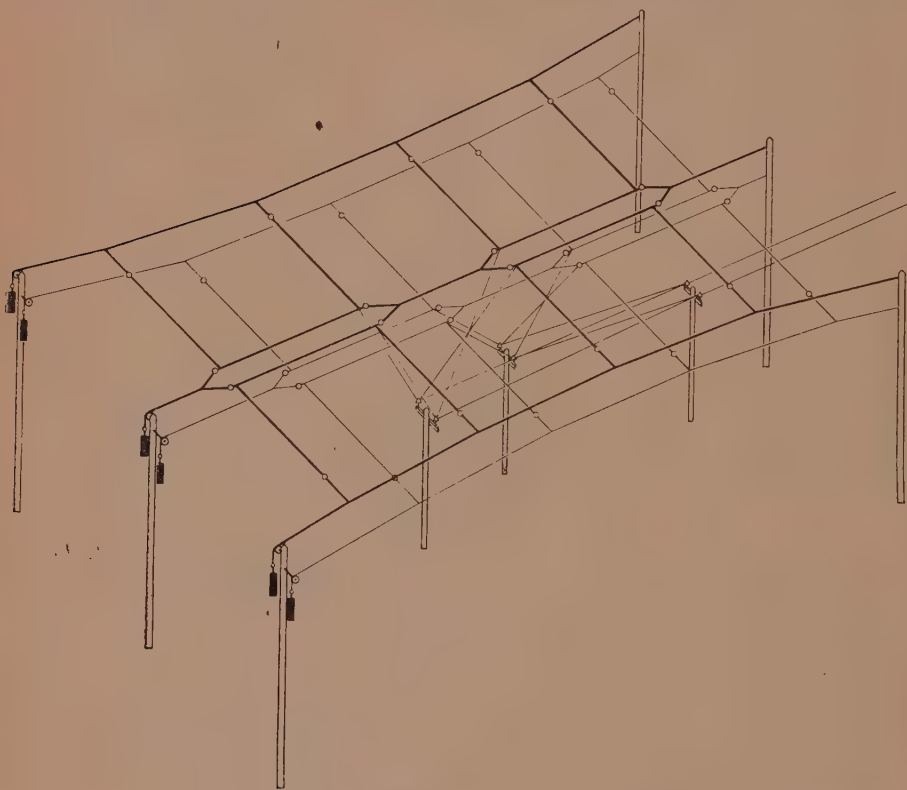


Rys. 8. Charakterystyka promieniowania w płaszczyźnie dipoli anteny wzdłużnej normalnej $j|8DD|1$ (w układzie współrzędnych biegunowych).

Wg S. Manczarskiego

a zysk kierunkowy, również w stosunku do dipola półfalowego, około 7,7. Wartość tego zysku związana jest z potrzebną dla celów radiofonicznych dość znaczną szerokością listka głównego (rys. 8). Umieszczenie anteny nad ziemią na wysokości nie mniejszej niż $\frac{\lambda}{2}$ w przybliżeniu nie

wpływa na oporność promieniowania anteny, natomiast maksymalny zysk kierunkowy ulega prawie czterokrotnemu zwiększeniu. Od wysokości anteny nad ziemią zależy również kąt elawacji maksimum głównego listka promieniowania; przy wysokości równej $\frac{\lambda}{2}$ wymieniony kąt elewacji wynosi około 25° , co odpowiada dobrze warunkom naświetlania terenów, leżących w najdalej położonych od nas krajach europejskich, przy wyko-



Rys. 9. Antena wzdłużna z falą stojąco-bieżącą.

Wg S. Manczarskiego

rzystaniu fal jonosferycznych odbijanych od warstwy $F2$. Dla obliczenia zysku anteny w odniesieniu do anteny izotropowej zysk względem dipola półfalowego należy pomnożyć przez współczynnik 1,64.

Dzięki znacznej oporności promieniowania antena $j|8DD|1$ może być obciążona bardzo dużą mocą, rzędu kilkuset kilowatów. Sprawność anteny jest bliska jedności.

Na rysunku 9 przedstawiona jest jedna z wersji praktycznego układu anteny wzdłużnej $j|8DD|1$. Dipole promieniujące utrzymywane są w dwóch blisko siebie położonych płaszczyznach (ściśle biorąc powierz-

chniach) za pomocą przeciwwag mechanicznych. Przesunięcie fazy o 90° pomiędzy prądami płynącymi w dwóch czwórkach dipoli osiąga się przez wytworzenie odpowiedniej różnicy dróg w poziomych fiderach, zasilających te czwórki.

Dla przejrzystości rysunku 9 pominięto tam niektóre szczegóły, jak na przykład dodatkowe druty przy dipolach promieniujących, mające na celu zmniejszenie oporności falowej tych dipoli oraz dodatkowe pętle, dopasowujące poziome fidera i służące do wytwarzania w nich fali bieżącej.

Oprócz zasadniczych zalet rozpatrywanej anteny, wynikających ze spełnienia postawionych dla niej warunków, należy dodać jeszcze następującą: w antenie danego typu stosunkowo łatwo daje się dokonać zmianę kierunku promieniowania anteny o 180° , co osiąga się przez proste odwrócenie kierunku prądów w jednej z czwórek dipoli promieniujących.

Anteny wzdłużne z falą stojąco-biejącą zostały wybudowane przez Polskie Radio i oddane do eksploatacji w latach 1950—1951. Przy budowie i pomiarach tych anten wziął udział oprócz autora — jako projektodawcy — zespół inżynierów Polskiego Radia w osobach: W. Rabęcki, H. Kalita, W. Lisicki i J. Ziółkowski. Przeprowadzone pomiary wykazały słuszność obliczeń matematycznych zawartych w projekcie.

WYKAZ LITERATURY

1. Ajzenberg G. Z.: *Antenny dla magistralnych radioswiazek*. 1948.
2. Dolph C. L.: *A current distribution for broadside arrays which optimizes the relationship between beam width and side lobe level*. Proc. I. R. E., vol. 34, nr 6, June 1946, pp. 335—348.
3. Duhamel R. H.: *Optimum patterns endfire arrays*. Proc. I. R. E., vol. 45, May 1953, pp. 652—659.
4. Goward F. K.: *An improvement in endfire arrays*. Proc. I. E. E., part III, vol. 94, November 1947, pp. 415—418.
5. Hansen W. W. and Woodyard J. R.: *A new principle in directional antennas design*. Proc. I. R. E., vol. 26, March 1938, pp. 333—345.
6. Manczarski S.: *Tablice do wykładów politechnicznych z anten*. Warszawa, 1950 r.
7. Manczarski S.: *Studium nad antenami wzdłużnymi z punktu widzenia potrzeb radiofonii krótkofalowej o zasięgu europejskim*, część I, II, III. Praca wykonana dla Polskiego Radia, Warszawa, 1950—1951.
8. Manczarski S.: *Rozpraszanie fal w jonosferze i jego praktyczne wykorzystanie (antena PBS)*. Kursokonferencja, Karpacz, 1954.
9. Schelkunoff S. A.: *A mathematical theory of linear arrays*. Bell Syst. Tech. Jour., January 1943, vol. 22, pp. 80—107.
10. Woodward P. M.: *A method of calculating the field over a plane aperture required to produce a given polar diagram*. Proc. I. E. E., part III A, vol. 93, March-May 1946, pp. 1554—1558.

С. МАНЧАРСКИ

УМЕНЬШЕНИЕ БОКОВЫХ ЛЕПЕСТКОВ В НАПРАВЛЕННЫХ АНТЕННАХ

Резюме

Уменьшение боковых лепестков в характеристиках излучения коротковолновых антенн радиосвязи соответствует выраженному CCLR стремлению ослабить взаимные возмущения передатчиков при помощи удаления или уменьшения излучений энергии в нежелательных направлениях.

В настоящей работе рассмотрены три основных типа коротковолновых направленных антенн радиосвязи:

- 1) ромбические антенны,
- 2) синфазные антенны,
- 3) антенны бегущей волны.

Ввиду их широкополосности теперь чаще всего применяются ромбические антенны, которые, однако, в сравнении с другими типами антенн отличаются сравнительно большими размерами боковых лепестков. Советская система двойного ромба Айзенберга — рис. 1, 2 и 3 — в некоторой степени уменьшает боковые лепестки. Как видно из этих рисунков, более значительное уменьшение боковых лепестков получается только вблизи некоторой определённой частоты, напр. на рис. 2 для $\lambda \approx 1.33 \lambda_0$, где λ_0 является оптимальной длиной волны ромба по отношению к излучению главного лепестка.

Относительно синфазных антенн — известным способом уменьшения их боковых лепестков является образование неравномерного распределения тока в элементах антенны.

Это распределение подбирается таким образом, чтобы наибольший ток получился в серединных элементах.

Простейший пример применения этого метода показан на рис. 4. На этом рисунке под диаграммой приведены формулы, облегчающие расчёт соответственных кривых. Числовые значения коэффициентов, входящих в состав приведенных формул, определяются по таблице на рис. 5. Эта таблица составлена автором для быстрого расчёта характеристик излучения как синфазных антенн, так и антенн бегущей волны.

При выступании многих боковых лепестков возникает вопрос, каким образом следует распределить ток в отдельных элементах антенны, чтобы достигнуть одновременного уменьшения уровня всех боковых лепестков. Теоретически вопрос этот был решён в 1946 году С. Л. Дольфом, доказавшим, что, применяя многочлен Чебышева, является возможной проектировка такого распределения тока в синфазной антенне, при которой уровень всех боковых лепестков составляет некоторое постоянное, заранее избранное значение. Многочлен Чебышева N -ного порядка определен выражениями (1) и (2).

На практике, при реализации метода Дольфа встречаются, однако, очень серьезные затруднения, в особенности при синфазных антеннах со значительным количеством элементов.

Не лучше обстоит этот вопрос и с антеннами бегущей волны. Как и для синфазных антенн, здесь тоже является теоретически возможным произвольно уменьшить боковые лепестки при помощи неравномерного распределения тока в антенне бегущей волны.

В 1953 году Дюгамель доказал, что благодаря применению многочлена Чебышева можно в антенне этого типа свести все боковые лепестки к некоторому постоянному, заранее избранному уровню. Пояснение такого решения мы видим на рис. 6.

Как следует из поданных на этом рисунке значений токов, метод Дюгамеля вызывает уменьшение сопротивления излучения антенны. Это обстоятельство, а также осложнения, связанные с регулировкой токов и их фазовым смещением создают очень значительные затруднения для практического применения метода Дюгамеля. Подобные осложнения выступают и при методе Щелкунова (супер-направленные антенны), а также при методе Говарда.

В Польше исследования передающих направленных антенн бегущей волны с уменьшенными боковыми лепестками были начаты в 1950 году. Polskie Radio, интересовавшееся этим вопросом, поставило для антенны бегущей волны следующие предварительные требования:

а) возможно наилучшая концентрация излучаемой энергии в требуемом угле, т.е. наиболее соответствующая форма главного лепестка и уменьшение боковых лепестков,

б) возможность нагрузки антенны значительной мощностью — до нескольких сот киловатт,

в) значительный коэффициент полезного действия антенны,

г) простое и дешёвое как схематическое, так и конструкционное решение, не требующее сложной регулировки.

Ввиду этих требований, решение разработанное автором, основано на нормальной антенне бегущей волны, имеющей достаточное сопротивление излучения, т.е. на антенне с т. наз. стационарно-бегущей волной. Применение такой волны являлось новостью в конструкции антенн бегущей волны со многими элементами; оно позволило удалить поглощающее энергию сопротивление, которое было всегда применяемо в антеннах бегущей волны. Благодаря этому удалось достигнуть высокого коэффициента полезного действия новой антенны.

Распределение тока и фазовые соотношения в антенне со стационарно-бегущей волной изъяснены на рис. 7 и 8, на которых показана антенна, обозначенная символом $j8DD|1$. На рисунках этих указана тоже в прямоугольных и полярных координатах диаграмма направленности, а также формулы, по которым могут быть произведены расчёты. Как следует из этих чертежей, боковые лепестки сведены к незначительному уровню, без нужды сложной регулировки амплитуды и фазы токов в отдельных диполях.

Благодаря довольно значительному сопротивлению излучения антенны $j8DD|1$ — она может быть нагружена очень большой мощностью, порядка нескольких сот киловатт.

Коэффициент полезного действия антенны близок к единице. Достоинством описанной антенны является также возможность сравнительно лёгкого изменения направления её излучения на 180° .

На рис. 9 представлен один из вариантов практической схемы антенны $j8DD|1$.

Описанные антенны со стационарно-бегущей волной применяются в Польше в коротковолновом радиовещании.

S. MANCZARSKI

REDUCTION OF SIDE LOBES IN DIRECTIONAL ANTENNAS

Summary

In order to minimize the interference of two or more shortwave transmitting stations, the CCIR has been urging the elimination or reduction of radiation in unrequired directions. The reduction of side lobes in short-wave antenna radiation patterns will decrease such interference. In this paper, three basic types of directional short-wave radio-communication antennas are reviewed:

- 1) rhombic antennas,
- 2) broadside arrays,
- 3) endfire arrays.

Because of their wide-band characteristics, the rhombic antennas are most frequently used. These antennas, however, produce considerably greater side lobe radiation than the other types. The Eisenberg double-rhombic system, developed in the Soviet Union, reduces the side lobes only to a certain extent. As can be seen from Figs. 1, 2 and 3, considerable side lobe reduction occurs only in the neighbourhood of one particular frequency. In Fig. 2, for example, it can be seen that this frequency corresponds to $\lambda \approx 1.33 \lambda_0$ where λ_0 is the optimum rhombic wave length for the main lobe radiation.

A well known method for reducing the side lobes in broadside arrays is to maintain an uneven distribution of currents in the individual radiators. The distribution is chosen so as to provide the greatest current in the central radiators. The most simple example of this method is shown in Fig. 4.

Formulae for computing the points on the curves are to be found in the same figure under the curves in question. The numerical values of the coefficients appearing in these formulae can be found in the table of Fig. 5. This table has been prepared by the author to facilitate rapid calculation of both broadside and endfire array radiation patterns.

In those cases in which there are many side lobes, the problem arises of how to determine the distribution of currents in the individual radiators in order to decrease simultaneously the radiation in each side lobe.

The theoretical aspect of this problem was solved in 1946 by C. L. Dolph. He showed that by using the Tchebycheff polynomials, it is possible to design the current distribution for the array so that the level of radiation in each side lobe will be equal to a predetermined value. Tchebycheff polynomials of the order N are defined by Eqs. (1) and (2).

In practice, however, Dolph's method meets with many serious difficulties, especially for arrays with a considerable number of radiators.

Attempts to reduce side lobes of endfire arrays have fared no better. Here too, it is theoretically possible to achieve any arbitrary reduction of the side lobe radiation by creating an uneven current distribution in the array. In 1953, R. H. Duhamel showed that by using the Tchebycheff polynomials, it is possible to maintain the level of side lobe radiation at a predetermined value. Fig. 6 illustrates this method.

As can be seen from the values of the currents shown in Fig 6, the Duhamel method reduces the array radiation resistance. This fact coupled with the complications involved in controlling the current and the current phase relations creates very serious difficulties for any practical application of Duhamel's method. Similar

difficulties appear in the methods proposed by Schelkunoff (super-directional antenna) and by Goward.

Work on reducing side lobes of endfire transmitting arrays has begun in Poland in 1950. Polish Radio, which is interested in this problem, set the following requirements for endfire arrays:

- a) the best possible concentration of radiated energy in a given angle; e. i., in the most suitable shape of the main lobe with minimum side lobes,
- b) possibility for loading the antenna to a power of several hundred kilowatts,
- c) high antenna efficiency,
- d) simple and inexpensive circuits and constructional features not requiring complicated regulation.

In face of these requirements, the author worked out a system based on the uniform endfire array with a sufficiently high radiation resistance, the so-called standing-travelling wave antenna. The use of such a wave is an innovation in the construction of multi-radiator endfire arrays. It avoids the necessity of energy-absorbing terminating resistances always used in travelling wave arrays. As a result, a highly efficient new type of array was obtained.

The current distribution and phase relations in the endfire array with a standing travelling wave is explained in Figs. 7 and 8, where the array is designated by the symbol $j8DD|1$. The radiation characteristic in rectangular and polar coordinates are given in these figures along with the formulae from which they can be derived.

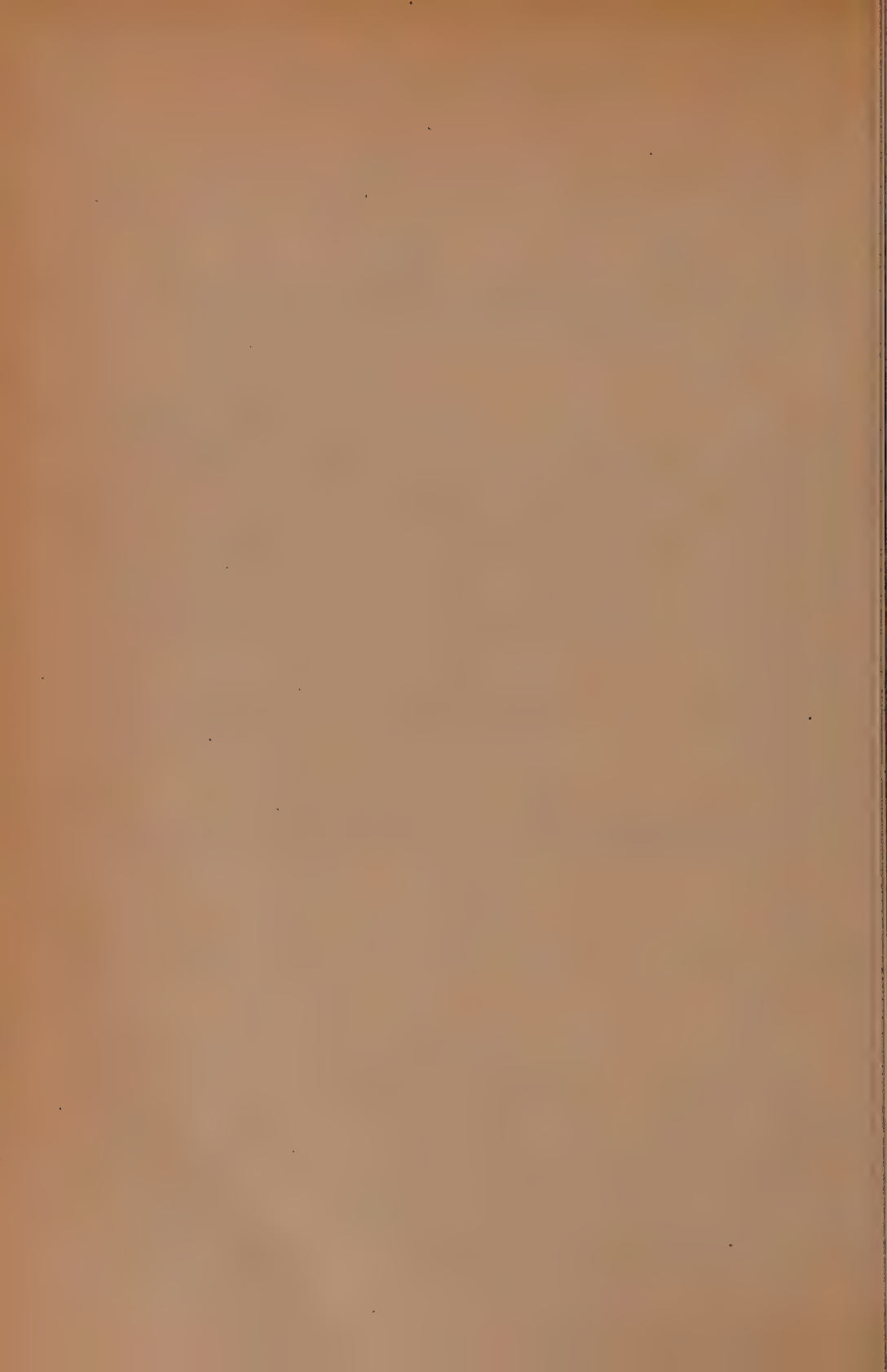
As can be seen from Fig. 7 and 8, the side lobes have been reduced to a negligible value without using complicated amplitude and phase regulation for the individual dipoles.

Because of its very high radiation resistance, the $j|8DD|1$ endfire array can be loaded to as much as several hundred kilowatts. The efficiency of the array is close to unity.

Another advantage of this array is that its direction can easily be changed by 180° .

Fig. 8. shows one practical version of the $j|8DD|1$ endfire array.

The described endfire array with a standing-travelling wave is being used in polish short-wave broadcasting.



621.396.11

S. BOROWSKI, S. JASIŃSKI, S. MANCZARSKI

Zjawisko Dopplera w propagacji jonosferycznej

Rękopis dostarczono 13. 12. 1955

Praca podaje zestawienie wyników pomiarowych odnośnie zjawiska Dopplera w oświetleniu klasycznej teorii tego zjawiska. Ponieważ niektóre wyniki doświadczalne zmian częstotliwości przypisywane zjawisku Dopplera znacznie odbiegają od obliczeń teoretycznych, przedstawiono przypuszczalny mechanizm zmian częstotliwości w warunkach granicznych (bliskich MUF) odbić od jonosfery.

1. WSTĘP

Celem pracy jest wyjaśnienie rozbieżności poglądów, jakie spotyka się w literaturze na temat znaczenia w radiokomunikacji krótkofalowej zjawiska Dopplera, które zachodzi w związku z niestabilnością jonosfery. Cytowane przez różnych autorów (Dodatek 1) pomiarowe wyniki wahań częstotliwości powodowanych przez to zjawisko różnią się między sobą o kilka rzędów wielkości. Niektórzy autorzy wiążą występowanie silnego efektu Dopplera z dużą zmiennością kąta przychodzenia fali jonosferycznej.

Według badań przeprowadzonych na falach ultrakrótkich (poniżej 150 MHz) w ostatnich latach w rejonach zorzy polarnej stwierdzono, że przy odbiciach od kotary zorzowej występują dość znaczne zmiany częstotliwości, które mogą być przypisane zjawisku Dopplera. Podczas jednej z serii prób przy 50 MHz otrzymano zmiany częstotliwości fali nośnej zawarte w granicach od 0 do 200 Hz [6].

2. KLASYCZNA TEORIA ZJAWISKA DOPPLERA NA TLE WYNIKÓW DOŚWIADCZALNYCH

Zjawisko Dopplera znane w akustyce i w optyce (przy zjawiskach astronomicznych) obserwowane jest również przy propagacji fal radiowych odbijanych od jonosfery.

Zjawisko to polega na tym, że częstotliwość odbierana różni się od nadawanej. Według klasycznej teorii zjawiska Dopplera obserwowana zmiana częstotliwości może wynikać na skutek:

1. ruchu źródła,
2. ruchu odbiornika,
3. ruchu środowiska (w przypadku propagacji jonosferycznej fal radiowych — skutek ruchu warstwy odbijającej),
4. kombinacji ruchów wyszczególnionych w punktach 1. i 3.

Zjawisko Dopplera przy ruchach zwierciadlanej warstwy jonosferycznej podane zostało w ujęciu optyki geometrycznej w Dodatku 2. Z dyskusji wzorów przytoczonych w tym Dodatku wynika, że zjawisko Dopplera przy propagacji jonosferycznej może być wywołane jedynie przez składową pionową szybkości poruszania się warstwy odbijającej. Zmiana częstotliwości (względna) wyraża się wówczas wzorem

$$\frac{\Delta f}{f} = 2 \frac{u}{c} \sin \alpha, \quad (1)$$

gdzie

Δf — zmiana częstotliwości (na skutek zjawiska Dopplera),

f — częstotliwość emitowana,

u — składowa pionowa szybkości unoszenia się warstwy,

c — szybkość światła,

α — kąt elewacji.

Ze wzoru (1) wynika, że $\frac{\Delta f}{f}$ jest tym większe, im większa jest szybkość u i kąt α .

W tablicy 1 podano wyniki obliczeń $\frac{\Delta f}{f}$ dla kątów elewacji 10° i 30° na podstawie wzoru (1) przy obserwowanej, na przykład w Anglii w okresach małych plam słonecznych, szybkości unoszenia się warstwy odbijającej $u \approx 12$ m/sek.

Tablica 1

α°	u m/sek	$\frac{\Delta f}{f} \cdot 10^8$
10	12	1,4
30	12	4

Z porównania powyższych obliczeń z danymi Dodatku 1 wynika, że w niektórych przypadkach jest dość duża zgodność teorii z obserwacją. Rezultaty obliczeń różnią się jednak znacznie od wyników opublikowanych przez Böhma w 1929 r. Pewne światło rzuca na powyższe rozbieżności ta okoliczność, że ujęcie optyki geometrycznej jest słuszne tylko dopóty, dopóki częstotliwość robocza nie stanie się zbyt bliska granicznej częstotliwości jonosfery; w tym momencie optyka geometryczna przestaje obowiązywać.

Sprawa ta wymaga dodatkowych wyjaśnień. Według praw optyki geometrycznej przy zmniejszaniu się gęstości jonizacji przejście od stanu odbijania się fali jonosferycznej do stanu nieodbijania się jej odbywa się skokowo. Graniczne odbicie zachodzi wtedy, gdy szybkość grupowa fali staje się równa zero. Ponieważ jednak iloczyn szybkości grupowej przez fazową równa się kwadratowi szybkości światła, to w miarę zbliżania się do częstotliwości granicznej (MUF), gdy szybkość grupowa dąży do zera, szybkość fazowa dąży do nieskończoności. Uwzględniając jeszcze, że jonosfera jest niejednorodna, należy zauważyć, że niejednorodności stają się wtedy małe w porównaniu z długością fali w danym miejscu. W ten sposób powstają właśnie warunki, w których optyka geometryczna przestaje obowiązywać wobec następującej ogólnej zasady: metody optyki geometrycznej stosują się tylko wtedy, gdy koncentracja elektronowa nie ulega praktycznie zmianie w granicach odległości porównywalnych z długością fali.

W rozpatrywanych warunkach musimy się zatem uciec do optyki falowej. W związku z tym niżej rozpatrzono mechanizm odbić jonosferycznych w ujęciu optyki falowej i w oparciu o zdjęte doświadczalnie przebiegi pojedynczych impulsów powracających na ziemię po odbiciu od jonosfery.

3. PRZYPUSZCZALNY MECHANIZM ODBIĆ OD JONOSFERY W WARUNKACH GRANICZNYCH

Ścisła interpretacja matematyczna zjawisk zachodzących w jonosferze natrafia w ujęciu optyki falowej na bardzo poważne trudności, zwłaszcza w pobliżu stanów granicznych.

Ogólne zagadnienie rozchodzenia się fal elektromagnetycznych w środowisku zjonizowanym znajdującym się w polu magnetycznym ziemi nie znalazło dotąd ścisłego rozwiązania. Jest to spowodowane tym, że:

1. nie zostały jeszcze wyprowadzone ścisłe równania rozchodzenia się fal w rozważanych warunkach,
2. przy rozwiązywaniu równań trudno jest sformułować w sposób właściwy warunki graniczne.

Tylko w niektórych przypadkach szczególnych rozkładu jonizacji i dla niektórych kierunków specjalnych lub wobec braku pola magnetycznego równania mogą być sformułowane i może być znalezione ich rozwiązanie. Dlatego spróbowano odtworzyć przypuszczalny mechanizm odbić jonosferycznych w tych warunkach w sposób przybliżony, opierając się na pewnych materiałach doświadczalnych. Jako punkt wyjścia rozpatrzono, stwierdzone eksperymentalnie, zjawisko rozbicia na szereg ciągów pojedynczego impulsu, odbitego od warstwy jonosferycznej.

Przykład takiego impulsu pokazany jest w Dodatku 3.

Mechanizm rozbijania impulsu na ciągi można ewentualnie wytłuma-
czyć kilkoma sposobami, a mianowicie:

1. rozbite impulsu następuje wskutek rozszczepiającego działania pola geomagnetycznego w jonosferze, wytwarzającego falę zwyczajną i nadzwyczajną,

2. rozbite impulsu następuje wskutek zmienności w czasie charakterystyki filtru dolnoprzepustowego, jakim jest jonosfera w pobliżu częstotliwości granicznej (o charakterystyce tego filtru będzie mowa niżej).

Rozbity na szereg ciągów impuls może być rozpatrzony za pomocą całki Fouriera. W przypadku dla przejrzystości mocno uproszczonym, traktując poszczególne ciągi jako odcinki sinusoidy, można z łatwością otrzymać na tej drodze rozkład spektralny pojedynczego ciągu. Przykład takiego rozkładu spektralnego podany jest w Dodatku 4. Jak wynika

z przedstawionego w Dodatku wykresu dla $n=3$, wartości $\frac{\omega}{\omega^0} \approx 1,2$ odpowiada wartość $\frac{S\omega_0}{2} \approx 3$, co stanowi około 0,64 wartości maksymalnej $\frac{S\omega_0}{2}$

Im ciągi fal są dłuższe, tym pasmo staje się węższe. Dla ciągów złożonych na przykład z $n=100$ fal amplituda fali odbitej o częstotliwości różniącej się o $\frac{1}{10n} = 0,1\%$ od częstotliwości emitowanej jest prawie równa amplitudzie fali odbitej o częstotliwości emitowanej; dla częstotliwości emitowanej 20 MHz stanowi to różnicę 20 kHz, co jest wartością dość znaczną.

Przy stanach bliskich granicznemu jonosfera zachowuje się selektywnie jak filtr dolnoprzepustowy w stosunku do promieni powracających na ziemię. Wynika to z falowego ujęcia zjawiska odbicia, co znajduje swój wyraz w charakterystycznym przebiegu współczynnika odbicia, który zobrażowany jest w Dodatku 5. W Dodatku tym, ze względu na trudności matematyczne, przedstawiono współczynnik odbicia w funkcji różnicy częstotliwości krytycznej i bieżącej jedynie dla transmisji prostopadłej (częstotliwość krytyczna jest szczególnym przypadkiem częstotliwości granicznej). Analogicznego przebiegu współczynnika odbicia należy spodziewać się dla transmisji ukośnej.

W ujęciu optyki geometrycznej mielibyśmy prostokątny przebieg współczynnika odbicia R , co nie odpowiada rzeczywistości.

Ze względu na ustawicznie zmieniające się właściwości filtru jonosferycznego i związane z tym stany przejściowe, rozkład spektralny impulsów powracających na ziemię ulega różnym zmianom, zwłaszcza przy kilkakrotnym odbiciu od jonosfery.

Jak z tego wynika, szerokie widma występują tylko w warunkach granicznych, które w praktyce spotyka się oczywiście znacznie rzadziej niż przeciętne. Dla doświadczalnego potwierdzenia tego faktu przeprowadzono

szereg obserwacji na radiostacji Instytutu Łączności oraz na radiostacji odbiorczej Ministerstwa Łączności w Grodzisku.

W Instytucie Łączności badania przeprowadzono za pomocą nasłuchów (z filtrem) i przy użyciu oscyloskopu. Na radiostacji odbiorczej w Grodzisku próby dokonane były również metodą nasłuchów, a oprócz tego przez obserwację regulowanej automatycznie heterodyny, podążającej za częstotliwością odbieranego sygnału.

Należy zauważyć, że opisany mechanizm znacznej zmiany częstotliwości wyjaśnia związek zachodzący między nim a dużą zmiennością przychodzenia fal jonosferycznych.

Reasumując powyższe rozważania możemy stwierdzić, że przy dokonywaniu pomiarów na podstawie odbioru częstotliwości wzorcowych odległych radiostacji krótkofalowych należy unikać zbliżania się do stanów granicznych jonosfery.

Instytut Łączności

WYKAZ LITERATURY

1. Alpert J. L., Ginzburg W. L., Feinberg E. L.: *Rasprostranienije radiowołn*. Moskwa 1953.
2. Both C. F., Gregory G.: *Effect of Doppler's principle on the comparison of standard frequencies over a transatlantic radio path*. P. O. Elec. Eng. Jour., 40 pt 4:153 — 8, Jan. 1948.
3. Böhm C.: *Mehrfachwege und Doppler-Effekt bei der Ausbreitung von kurzen Wellen*. Telefunkenzeitung 10, nr 53 (1929) 9.
4. Griffiths H. V.: *Doppler effect in propagation*. Wireless Engineer nr 285, June 1947.
5. Manczarski S.: Wykłady politechniczne, 1954.
6. Moore R. K.: *A VHF propagation phenomenon associated with aurora*. Journ. Geophys. Res., vol. 56, March. 1951.
7. Rydbeck O.: *The reflection of electromagnetic waves from a parabolic ionized layer*. Phil. Mag., 34, 1943.
8. Rydbeck O.: *On the propagation of radio waves*. Trans. Chalmers University, Gethenborg, Sweden. nr 3, 1944.

Dodatek 1

Dane z literatury odnośnie zjawiska Dopplera

Źródło	Data obserwacji	Trasa i jej długość	Częstotliwość w Mc/s	$\left \frac{\Delta f}{f} \cdot 10^8 \right $	U w a g i
Böhm [3]	1929 r.	Niemcy-Buenos Aires, ok. 12 000 km	20	100 000	Z obserwacji fototelegrafii
Both i Gregory [2]	1934 r.	Anglia-Ameryka Północna, ok. 5050 km	5	20	Dane z obserwacji wzorca częstotliwości, dokładność wzorca $\frac{\Delta f}{f} \cdot 10^8 = \pm 5.$
	Grudzień 1946 r.	Anglia-Ameryka Północna, ok. 5050 km	10	średnio 4,7 107 63 7	Dokładność wzorca w ciągu paru dni $\frac{\Delta f}{f} \cdot 10^9 = \pm \text{kilka}.$
	Godz. 9—9.15 Godz. 20.42 Godz. 9.45				Zmiany częstotliwości największe o wschodzie i zachodzie słońca. Przy zachodzie słońca częstotliwość maleje, przy wschodzie rośnie.
	28. 4. 1947— —2. 5. 1947 r. Godz. 20—8	Anglia-Ameryka Północna, ok. 5050 km	10 15	do 20 10	Zmiany częstotliwości większe dla niższej częstotliwości.
Griffiths [4]	Listopad 1943 r.	Anglia-Ameryka Północna, ok. 5050 km	15	2—7 średnio 3	Dokładność częstotliwości stacji nadawczej $\frac{\Delta f}{f} \cdot 10^8 = \pm 1$
	Godz. 1 Grudzień 1945 r.		15	3,3	

Dodatek 2

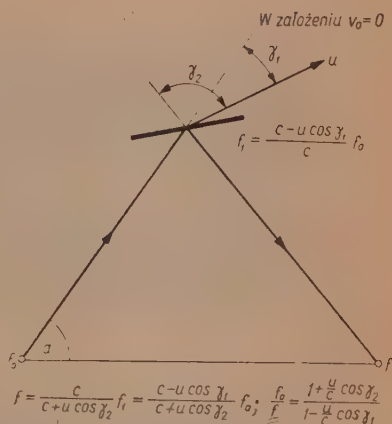
**Zjawisko Dopplera przy ruchach zwierciadlanej warstwy jonosferycznej
w ujęciu optyki geometrycznej**

Ogólny wzór Dopplera

$$f_s = \frac{c + v_0 - v_s}{c + v_0 - v_z} f_z$$

gdzie:

- f_s — częstotliwość odbierana przez słuchacza s
- f_z — częstotliwość wysyłana przez źródło z
- c — prędkość rozchodzenia się fal w środowisku
- v_0 — rzut prędkości środowiska na kierunek zs
- v_s — rzut prędkości słuchacza na kierunek zs
- v_z — rzut prędkości źródła na kierunek zs

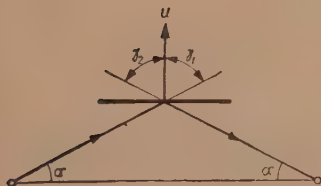


Dla $u \ll c$

$$\frac{f_0}{f} \approx \left(1 + \frac{u}{c} \cos \gamma_1\right) \left(1 + \frac{u}{c} \cos \gamma_2\right) \approx 1 + \frac{u}{c} (\cos \gamma_1 + \cos \gamma_2)$$

Przypadki szczególne

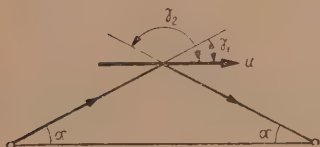
1. $\gamma_1 = \gamma_2 = 90^\circ - a$ $\frac{f_0}{f} = 1 + 2 \frac{u}{c} \sin a$



a) przy $a \rightarrow 0$; $\frac{f_0}{f} \rightarrow 1$

b) przy $a \rightarrow 90^\circ$; $\frac{f_0}{f} \rightarrow 1 + 2 \frac{u}{c}$; $\frac{\Delta f}{f} \rightarrow 2 \frac{u}{c}$

2.

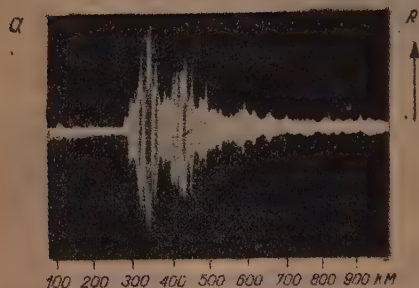


$$\left. \begin{array}{l} \gamma_1 = a \\ \gamma_2 = 180^\circ - a \end{array} \right\} \frac{f_0}{f} \approx 1 + \frac{u}{c} (\cos a - \cos a) = 1$$

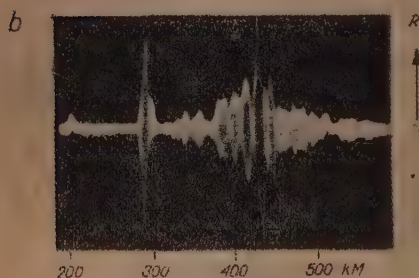
Wg Manczarskiego [5].

Dodatek 3

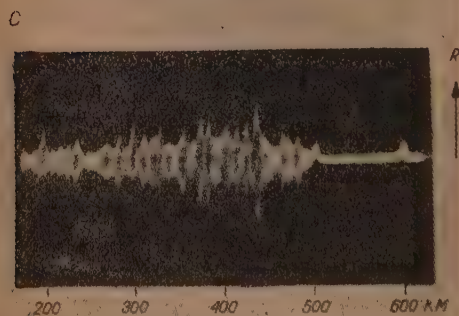
Zjawisko rozbicia na szereg ciągów pojedynczego impulsu odbitego od warstwy F_2



a. Rozbity i rozszerzony sygnał fali zwyczajnej; największy z ciągów zajmuje szerokość ok. 100 km (400—300 km), inne mniejsze ciągi mają szerokość podobnego rzędu.



b. Rozbity i rozszerzony sygnał fali zwyczajnej, zajmujący szerokość ok. 220 km (550—330 km); ciąg po lewej stronie odnosi się do fali nadzwyczajnej.

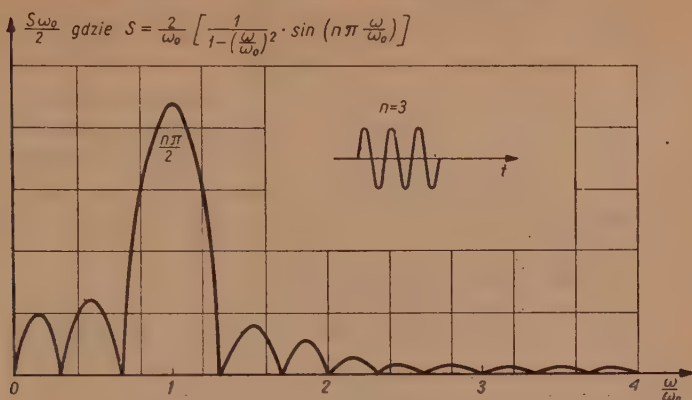


c. Zespół ciągów fali zwyczajnej i nadzwyczajnej, zajmujący łącznie szerokość ok. 300 km (500—200 km).

Według zdjęć fotograficznych, podanych przez Alperta w książce "Rasprostranienije radiowołn", 1953 [1].

Dodatek 4

Rozkład spektralny odcinka sinusoidy



S — moduł gęstości spektralnej, *

ω — pulsacja dowolnej składowej w paśmie,

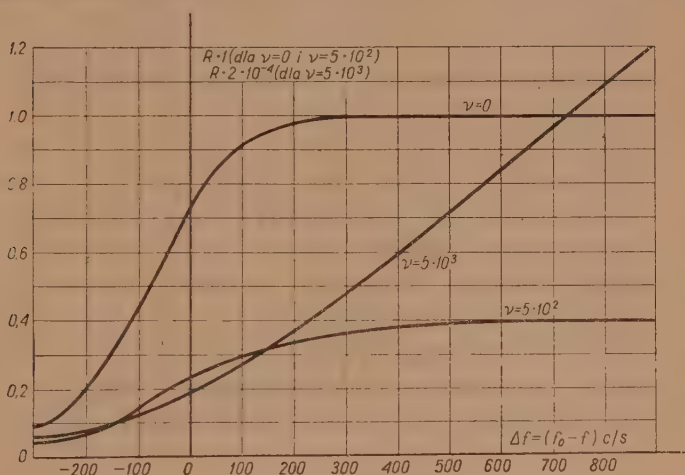
ω_0 — pulsacja częstotliwości emitowanej,

n — ilość fal sinusoidy.

* $\frac{S\omega_0}{2} = 0$ najbliżej max. dla $\left| 1 - \frac{\omega}{\omega_0} \right| = \frac{1}{n}$

Dodatek 5

Obliczony współczynnik odbicia od warstwy F w ujęciu optyki falowej



R — współczynnik odbicia,

ν — ilość zderzeń na sekundę elektronów z molekułami,

f_0 — częstotliwość krytyczna,

f — częstotliwość fali.

Obliczenia odpowiadają założeniom:

a) paraboliczny rozkład gęstości jonizacji w warstwie,

b) połowiczna grubość warstwy $z_m = 120$ km,

c) $f_0 = 10$ Mc/s.

Wg Ginzburga, "Rasprostranienije radiowołn", 1953 [1], [7].

С. БОРОВСКИ, С. ЯСИНЬСКИ, С. МАНЧАРСКИ

ЯВЛЕНИЕ ДОППЛЕРА В ИОНОСФЕРНОМ РАСПРОСТРАНЕНИИ РАДИОВОЛН

Резюме

Целью настоящей работы является изъяснение встречаемых в литературе воззрений на тему происходящего, в связи с нестабильностью ионосферы, явления Допплера и его значения в коротковолновой радиосвязи. Цитированные разными авторами результаты измерений колебания частоты, вызываемого этим явлением (приложение 1), отличаются друг от друга на несколько порядков.

Некоторые авторы видят причины возникновения сильного явления Допплера в большой изменчивости угла прихода ионосферной волны.

Доклад содержит как математический анализ проблемы (приложение 2), так и результаты измерений и наблюдений. Математическая формула колебаний частоты вследствие подвижности ионосферы (приложение 2), выведенная на основах геометрической оптики, даёт, как это следует из вычислений, незначительные изменения частоты. Это, однако, правильно только до момента, когда рабочая частота не станет слишком близкой к предельной частоте ионосферы МПЧ. В этот момент прекращается применение геометрической оптики. При состояниях близких к предельному избирательное поведение ионосферы подобно фильтру нижних частот по отношению к возвращающимся на землю лучам, причём свойства этого фильтра постоянно изменяются. В следствие возникающих таким образом переходных состояний возвращающийся на землю, отдельный импульс подвергается расщеплению на последовательные короткие группы (приложение 3). Если затем эти группы рассматривать при помощи интеграла Фурье (приложение 4), то получается довольно широкий спектр частот, приблизительно соответствующий данным некоторых авторов.

Распределение частот в спектре возвращающегося на землю сигнала может быть различно (прил. 4, 5). Из этого следует, что широкие спектры выступают только в предельных условиях, которые на практике встречаются очевидно гораздо реже чем средние. Для опытного подтверждения этого факта был исполнен ряд наблюдений радиостанцией Института Связи, а также приёмной станцией Министрства Связи в Гродзиске. Следует заметить, что описанный механизм сильного явления Допплера хорошо изъясняет связь между этим явлением и значительной изменчивостью угла прихода ионосферных волн.

Вытекающий из настоящей работы практический вывод тот, что при исполнении измерений на основании нормальных частот отдалённых коротковолновых станций — следует избегать приближения к предельным состояниям ионосферы.

S. BOROWSKI S. JASIŃSKI, S. MANCZARSKI

DOPPLER EFFECT IN IONOSPHERIC PROPAGATION

Summary

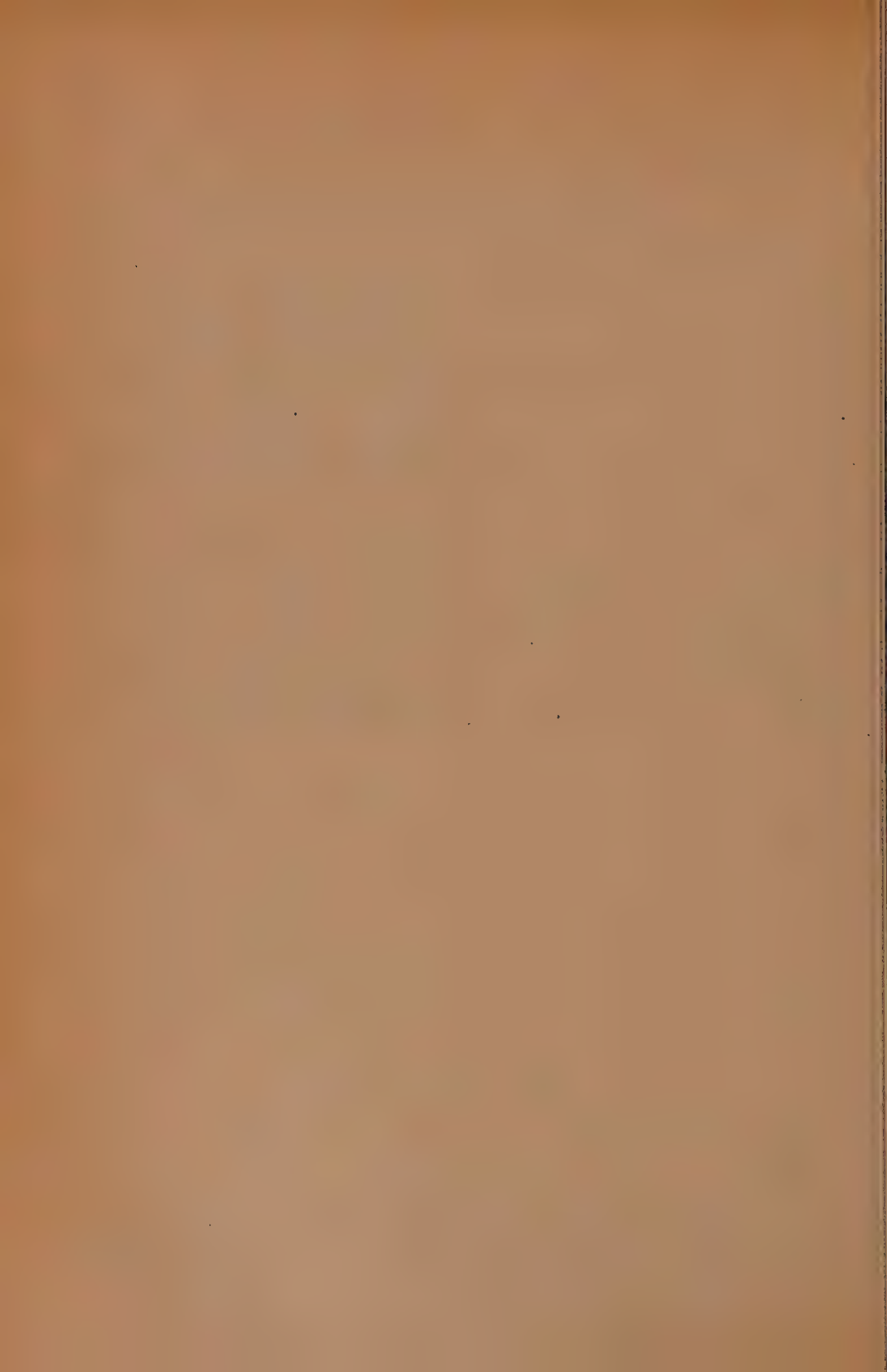
The object of this paper is to explain the difference in views found in the literature on the importance of the Doppler effect in short-wave communications, this effect arising from the variations of the ionosphere. Measurements of frequency variations due to Doppler effect, given by various authors, vary widely, even to the extent

of several orders (Appendix 1). Some authors associate the appearance of a strong Doppler effect with the great variability of the angle of incidence of the waves on the ionosphere.

This paper contains a mathematical analysis of the general problem (Appendix 2) as well as an analysis of the measurements and observations of the effect. A formula expressing the variation of frequency caused by the motion of the ionosphere (Appendix 2) has been derived on the basis of geometrical optics. Calculations made using this formula give a small change of frequency. The formula can be used as long as the operating frequency is not too close to the ionospheric MUF frequency. In this case geometrical optics is no longer applicable. Near this limit the ionosphere behaves like a selective low-pass filter for the waves returning to the earth, whereby the parameters of this filter are constantly changing. As a result of the changes arising in this manner, a single impulse returning to the earth is broken up into a series of short wave trains (Appendix 3). If these wave trains are studied with the aid of the Fourier integral (Appendix 4), we obtain a wide frequency spectrum, which more or less is in agreement with the data obtained by some authors. The spectral frequency distribution for the signal returning to the earth can vary widely (Appendix 4 and 5).

Analysis shows that the wide spectra appear only in the limiting conditions, which certainly occur considerably less often than the average ones. To confirm this fact experimentally, a series of observations were made at Warsaw and Grodzisk (Poland). It should be noted that the treatment presented by the authors provides a good explication of the correlation between the Doppler effect and the great variability of the angle of arrival of ionospheric waves.

As a practical conclusion results, it may be noted, that in carrying out measurements based on the reception of standard frequencies from distant short-wave radiostations, an approach to the limiting conditions should be avoided.



523.74

K. URBANIK, A. ZIEBA

Prognoza aktywności słońca

Rękopis dostarczono 18. 11. 1955

Zmiany elektromagnetycznej aktywności słońca, powodujące w szczególności zaburzenia w odbiorze radiowym, charakteryzuje się na ogół liczbami Wolfa, które opisują rozmieszczenie i ilość plam słonecznych. Ponieważ ściśła zależność liczb Wolfa od czasu nie jest znana, przyjmuje się, że tworzą one proces stochastyczny. W artykule omówione są krótko probabilistyczne metody prognozy liczb Wolfa: liniowa metoda Kołmogorowa-Wienera oraz metoda uwzględniająca koszt eksperymentu oparta na teorii funkcji decyzyjnych.

Aktywnością elektromagnetyczną słońca interesują się zarówno astronomowie i fizycy jak i radiotechnicy. Jej znaczenie dla radiotechniki polega na tym, że w dużym stopniu od niej zależy działanie służb telekomunikacyjnych. Zwiększenie aktywności wpływa ujemnie na jakość odbioru.

Aktywność słońca jest ściśle związana z występowaniem plam słonecznych. Dlatego opisujemy ją za pomocą tak zwanych liczb Wolfa, charakteryzujących ilość i rozmieszczenie plam na słońcu

$$r = k (10g + f)$$

r — jest to liczbą Wolfa,

g — oznacza ilość grup plam słonecznych (izolowane plamy uważamy za oddzielne grupy),

f — całkowitą ilość plam,

k — jest współczynnikiem zależnym od przyrządu pomiarowego, w przybliżeniu równym jedności.

Zależność liczb Wolfa od czasu nie jest dokładnie znana.

Obserwacje wskazują na okresowe nasilanie się liczby plam słonecznych, przy czym wyraźny jest tu cykl jedenastoletni.

W pierwszym przybliżeniu przyjmuje się, że liczby Wolfa zależą od czasu losowo. Pozwala to na wykorzystanie aparatu teorii procesów stochastycznych do badania aktywności słońca, a w szczególności do ustalania prognozy aktywności słońca.

Niech $r(t)$ będąca funkcją losową oznacza liczbę Wolfa odpowiadającą chwili t . Obok funkcji $r(t)$ wygodnie jest posługiwać się odchyleniem $\varrho(t)$ wielkości $r(t)$ od średniej wartości $\bar{r}(t)$ ¹

$$\varrho(t) = r(t) - \bar{r}(t).$$

Dla odchylenia $\varrho(t)$, które również jest funkcją losową, mamy

$$\bar{\varrho}(t) = 0.$$

Przyjmujemy także, zgodnie z doświadczeniem, że korelacja odchylen $\varrho(t)$ i $\varrho(t+\tau)$ zależy wyłącznie od długości przedziału czasowego τ , a więc funkcja korelacyjna ma postać

$$B_r(\tau) = \overline{\varrho(t+\tau)\varrho(t)}.$$

Odchylenia $\varrho(t)$ tworzą zatem stacjonarny proces stochastyczny. Przyjmemy, że znamy liczby Wolfa w chwilach $t_1, \dots, t_n$. Wówczas szacując $\bar{r}(t)$ średnią arytmetyczną

$$\bar{r}(t) = \frac{1}{n} \sum_{i=1}^n r(t_i),$$

można obliczyć odchylenia $\varrho(t_1), \dots, \varrho(t_n)$. Zagadnienie prognozy liczby Wolfa w chwili t , przy znanych liczbach Wolfa w chwilach $t_1, \dots, t_n$ sprowadza się więc do prognozy odchylenia $\varrho(t)$, gdy znane są odchylenia $\varrho(t_1), \dots, \varrho(t_n)$. Ograniczymy się do prognozy polegającej na oszacowaniu nieznanej wartości $\varrho(t)$ kombinacją liniową wartości $\varrho(t_1), \dots, \varrho(t_n)$. Poza funkcją korelacyjną w prognozie nie wykorzystujemy innych charakterystyk procesu stochastycznego $\varrho(t)$.

Teoria prognozy liniowej opracowana została przez A. N. Kołmogorowa, N. Wienera i innych.

Prognoza liniowa sprowadza się do znalezienia układu współczynników liczbowych $a_1, \dots, a_n$; odchylenie $\varrho(t)$ szacuje się wówczas kombinacją liniową $a_1\varrho(t_1) + \dots + a_n\varrho(t_n)$. Prognozę opisaną współczynnikami $a_1, \dots, a_n$ nazywamy optymalną, jeśli średni błąd kwadratowy prognozy jest najmniejszy, to znaczy

$$\sigma^2 = \overline{\left| \varrho(t) - \sum_{i=1}^n a_i \varrho(t_i) \right|^2} = \min_{a_1, \dots, a_n} \overline{\left| \varrho(t) - \sum_{i=1}^n a_i \varrho(t_i) \right|^2}.$$

Dowodzi się, że współczynniki $a_1, \dots, a_n$ opisujące najlepszą prognozę spełniają układ równań (por. [2])

$$\sum_{i=1}^n a_i B(t_i - t_k) = B(t - t_k) \quad (k = 1, 2, \dots, n) \quad (1)$$

¹ Kreska nad symbolem zmiennej losowej oznacza wartość średnią (oczekiwaną)

a średni błąd kwadratowy σ^2 optymalnej prognozy jest równy

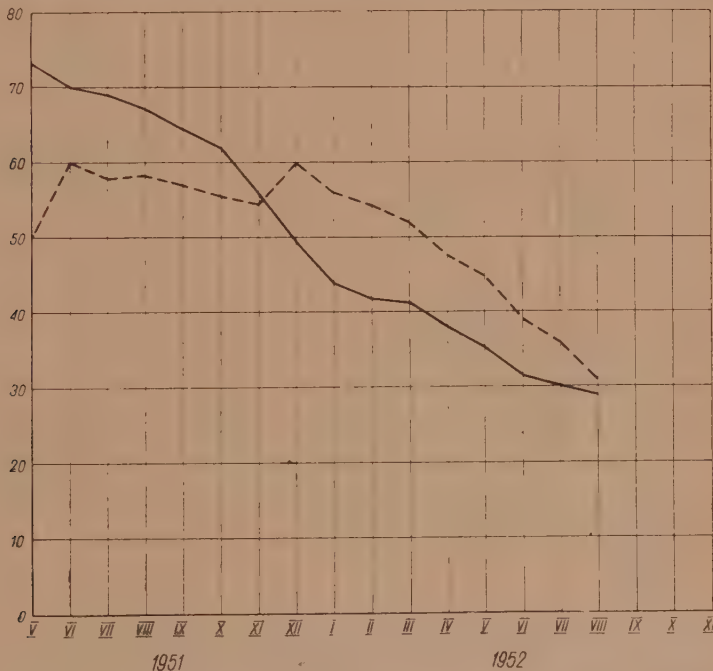
$$\sigma^2 = B(0) - \sum_{k=1}^n \sum_{i=1}^n a_k a_i B(t_i - t_k).$$

Odchylenia $\varrho(t_1), \dots, \varrho(t_n)$ możemy, jak powiedzieliśmy, obliczyć. Pozwala to oczywiście szacować funkcję korelacyjną $B(\tau)$ według znanego wzoru (por. [2] lub [6])

$$B(\tau) \approx \frac{1}{t_n - t_1 - \tau} \sum_{t_i + \tau < t_n} \varrho(t_i + \tau) \varrho(t_i).$$

Znając zaś funkcję $B(\tau)$ z równań (1) możemy wyznaczyć współczynniki optymalnej prognozy.

Zanim podamy pewne konkretne przykłady prognozy, zwróćmy uwagę, że omawiana teoria prognozy liniowej podaje rachunkowy sposób otrzy-



Rys. 1. Wykres aktywności słońca według prognozy i na podstawie obserwacji w latach 1951/1952.

mania optymalnej prognozy (w podanym wyżej sensie) na podstawie danych obserwacyjnych. Materiał obserwacyjny aktywności słońca jest niezmiernie bogaty i nie może być cały wykorzystywany w rachunkach liczbowych. Powstaje zatem zagadnienie wyboru optymalnej ilości obserwacji i ich optymalnego rozmieszczenia. Innymi słowy

powstaje pytanie — ile i które obserwacje należy wykorzystać. Zagadnienie to jest nierozwiązane. Próba, podjęta przez autorów tej pracy, wskazuje, że zagadnienie prowadzi do równań, których nawet bardzo niedokładne rozwiązanie jest praktycznie na razie niemożliwe. (Do zagadnienia tego powrócimy jeszcze przy omawianiu innych teorii prognozy). W teorii prognozy liniowej przyjmuje się heurystycznie także (por. [3], [4]) wzory typu

$$\varrho(t) = a_1 \varrho(t-1) + a_2 \varrho(t-2) + a_{10} \varrho(t-10) + a_{11} \varrho(t-11) + \\ + a_{20} \varrho(t-20) + a_{21} \varrho(t-21) + \dots$$

Na przykład, dla dwóch obserwacji otrzymuje się na $r(t)$ wzór

$$r(t) = 14,5 + 1,3 r(t-1) - 0,6 r(t-2).$$

Prognozę na lata 1925—1929 opartą na obserwacjach z dwóch poprzednich lat przedstawia poniższa tablica (por. [3]).

Rok	Prognoza	Obserwacja	Błąd
1925	33,2	44,3	11,1
1926	63,3	63,9	0,6
1927	71,5	69,0	-2,5
1928	65,4	77,8	12,4
1929	74,0	65,0	-9,0

Uwzględnijmy teraz tzw. koszt wykorzystania danych obserwacyjnych $C(N)$. Koszt ten należy rozumieć bardzo ogólnie; wchodzi do niego zarówno czas potrzebny do obliczeń jak i zużycie maszyn do liczenia itp. Oczywiście zależy na zmniejszeniu tego kosztu. Niech ϱ_N oznacza prognozę odchylenia $\varrho(t)$ opartą na N obserwacjach (obserwacje te traktujemy jako zmienne losowe, a więc ϱ_N jest wielkością losową). W (ϱ, ϱ_N) oznacza dokładność prognozy odchylenia liczby Wolfa od średniej wartości. Może to być na przykład

$$W(\varrho, \varrho_N) = |\varrho - \varrho_N|.$$

Zależy również na zmniejszeniu funkcji $W(\varrho, \varrho_N)$, to jest na polepszeniu dokładności prognozy.

Gdybyśmy umieli oszacować stratę gospodarczą wynikającą z danej błędnej prognozy, wówczas można by otrzymać nowe wyniki w zakresie poruszonego już zagadnienia ilości i rozmieszczenia w czasie danych obserwacyjnych. Dokładne oszacowanie tej straty jest przypuszczalnie trudne. Jest to oczywiście trudność pozamatematyczna. Ponieważ jednak w praktyce wystarczy oszacowanie przybliżone, podamy przykład wyznaczania ilości obserwacji w przypadku, gdy funkcja $C(N)$ i $W(\varrho, \varrho_N)$ są zna-

nymi funkcjami, których wartości wyrażają się w złotych. Wskażemy zresztą później pewne warianty tej teorii oparte na innych (w pewnych przypadkach dogodniejszych) założeniach o funkcji $W(\varrho, \varrho_N)$.

Przejdziemy teraz do teorii mających na celu prognozę dystrybuanty procesu. Rozpocznijmy od wąskiego zagadnienia szacowania ilości obserwacji. Później zajmiemy się zagadnieniami o wiele ogólniejszymi.

Sumę

$$R_e(N, \varrho_N) = C(N) + W(\varrho, \varrho_N)$$

nazywamy całkowitym kosztem prognozy, a wyrażenie

$$R(N, \varrho_N) = \max_{\varrho} [C(N) + W(\varrho, \varrho_N)]$$

maksymalnym całkowitym kosztem prognozy. Ilość N_0 obserwacji potrzebnych do prognozy nazywamy optymalną, jeżeli

$$R(N_0, \varrho_{N_0}) = \min_N R(N, \varrho_N).$$

Na prostym, szczególnym przykładzie pokażemy metodę znajdowania optymalnej ilości obserwacji.

Przyjmijmy najpierw pewne założenia upraszczające. Przede wszystkim zakładamy, że obserwacje wykonujemy w sposób ciągły, a nie dyskretny. Dysponujemy więc wykresem rejestrowanym przez aparat od pewnej chwili początkowej 0. Dalej zakładamy, że odchylenie $\varrho(t)$ liczb Wolfa od średniej tworzą proces normalny, czyli że rozkłady prawdopodobieństw są dane wzorem

$$P[\varrho(t) < x] = \frac{1}{\sqrt{2\pi t} \sigma} \int_{-\infty}^x e^{-\frac{(y-\mu t)^2}{2\sigma^2 t}} dy$$

i że prognoza dotyczy tylko niewiadomego parametru μ w tym rozkładzie. Przyjmijmy, że dokładność $W(\mu, \lambda)$ prognozy jest funkcją postaci

$$W(\mu, \lambda) = w(|\mu - \lambda|),$$

gdzie

μ — jest prawdziwą,

λ — przepowiedzianą wartością parametru.

Prognoza wielkości μ polega na oszacowaniu za pomocą średniej arytmetycznej wyników obserwacyjnych.

Dowodzi się [1], że optymalna ilość T pomiarów przy pewnych założeniach regularności funkcji $w(x)$ i $C(x)$ (które w praktyce można przyjąć) otrzymuje się z rozwiązania równania

$$-\frac{dC(T)}{dT} + \frac{\int_0^\infty w(x) e^{-\frac{x^2 T}{2\sigma^2}} dx}{\sqrt{2\pi T} \cdot \sigma} - \frac{\sqrt{2T}}{\sqrt{\pi} \sigma} \int_0^\infty \frac{w(x) x^2}{2\sigma^2} e^{-\frac{x^2 T}{2\sigma^2}} dx = 0$$

względem T . Znając funkcje $C(T)$ i $w(x)$ można to równanie rozwiązać metodami przybliżonymi.

Podamy teraz pewien inny typ prognozy, mający o wiele szerszy zasięg niż podana na przykładzie teoria wyznaczania optymalnej obserwacji. Teoria ta, oprócz optymalnej ilości i optymalnego rozmieszczenia pomiarów, pozwala na skonstruowanie optymalnej prognozy [5].

Podstawowym pojęciem tej teorii jest pojęcie funkcji decyzyjnej.

Przez funkcję decyzyjną rozumiemy ciąg przepisów, z których pierwszy orzeka, jakie obserwacje powinniśmy wykonać i uwzględnić na początku, drugi — na podstawie tych wyników — bądź orzeka, że wykonujemy i uwzględniamy dalsze obserwacje (i jeśli tak, to jakie), bądź wydaje tzw. decyzję ostateczną. Trzeci przepis wchodzi w rachubę tylko wtedy, jeśli drugi nie wydał decyzji ostatecznej i znowu, na podstawie wyników obserwacji dokonanych na obu poprzednich etapach bądź zaleca dalsze pomiary, bądź wydaje decyzję ostateczną itd. Przez decyzję ostateczną rozumiemy orzeczenie, że badany proces stochastyczny ma taką a taką dystrybuantę. W samym pojęciu funkcji decyzyjnej abstrahujemy od tego czy jej zalecenia są dobre, czy decyzja ostateczna jest prawdziwa itp. Przez dystrybuantę procesu rozumiemy funkcję

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = P\{\varrho(t_1) < x_1, \dots, \varrho(t_n) < x_n\}$$

określaną dla wszystkich układów $t_1, \dots, t_n, x_1, \dots, x_n$, gdzie symbol $P\{\}$ oznacza, jak zwykle, prawdopodobieństwo zdarzenia opisanego w nawiasie.

Używane w teorii funkcji decyzyjnych badanie etapowe nazywamy postępowaniem sekwencyjnym.

Oprócz wyżej opisanych funkcji decyzyjnych używamy funkcji decyzyjnych zrandomizowanych (losowych) [5]. Ma to zastosowanie do przypadków, gdy prognozę konstruujemy nie raz, lecz wiele razy. Możemy wtedy w poszczególnych badaniach stosować różne funkcje decyzyjne. Jeśli, ogólnie, z danego układu funkcji decyzyjnych za pomocą ustalonego mechanizmu (kostki) będziemy losowali poszczególne funkcje i stosowali je do kolejnych prognoz, będziemy mówili, że posługujemy się funkcją decyzyjną zrandomizowaną.

Zalety randomizacji można stwierdzić na prostym, nawet nie sekwencyjnym przykładzie. Ktoś kładzie monetę, my dajemy prognozę pojawienia się orła. Zadanie powtarza się wiele razy. Jeśli będziemy stale podejmowali jedną i tę samą decyzję z dwu możliwych: pojawi się orzeł, pojawi się reszka, to, jeżeli moneta była rzucana przypadkowo, możemy gwarantować, że średnio co najmniej (a także co najwyżej) połowa prognoz będzie dobra. Jeśli posłużymy się decyzją zrandomizowaną i sami swoje odpowiedzi: orzeł, reszka, będziemy losowali monetą, to

także będziemy mogli zagwarantować, że średnio połowa naszych prognoz będzie prawdziwa, ale już nie będziemy potrzebowali zakładać jak poprzednio, że moneta, której rzuty mamy odgadywać, jest rzucona przypadkowo. Nie będziemy potrzebowali w ogóle żadnych hipotez na ten temat. Randomizacja ma jeszcze inne zalety, których nie będziemy tu szczegółowo rozważali.

Przypuśćmy, że mamy daną rodzinę kompletnych dystrybuant procesów stochastycznych. Dystrybuanty oznaczmy symbolami $F_1, \dots, F_n$. Tę daną rodzinę będziemy nazywali bazą. Utwórzmy zrandomizowaną kombinację liniową

$$a_1 F_1 + a_2 F_2 + \dots + a_n F_n \quad (a_i \geq 0, \sum a_i = 1) \quad (2)$$

bazy. Dzięki normalizacji, taką kombinację znów możemy uważać za dystrybuantę.

Przypuśćmy, że tak wybraliśmy bazę dystrybuant (w praktyce kilkadziesiąt), iż mamy prawo przypuszczać, że losowy proces aktywności słońca daje się w przybliżeniu opisać jakąś z dystrybuant postaci (2). Przypuśćmy dalej, że wybraliśmy układ funkcji decyzyjnych $d_1, \dots, d_m$ (w praktyce — kilkadziesiąt sekwencyjnych metod testowania hipotez, że proces $\varrho(t)$ ma dystrybuantę $F_1, F_2, \dots, F_n$. Na ich podstawie skonstruujemy optymalną funkcję decyzyjną do badania procesów kształtu (2).

Niech $C(N)$ oznacza, jak poprzednio, koszt wykorzystania N obserwacji, $W(F_i, F_j)$ — stratę gospodarczą wynikającą z powzięcia decyzji ostatecznej orzekającej, że proces $\varrho(t)$ opisany jest dystrybuantą F_j , podczas gdy naprawdę opisany jest dystrybuantą F_i [$W(F_i, F_j) \geq 0$, $W(F_i, F_i) = 0$]. Przypuśćmy na chwilę, że proces aktywności słońca $\varrho(t)$ opisany jest dystrybuantą F_i i że posługujemy się funkcją decyzyjną d_k . Znając dystrybuantę F_i i funkcję decyzyjną d_k łatwo możemy obliczyć średni koszt całkowity² badania za pomocą funkcji d_k , procesu opisanego dystrybuantą F_i , koszt ten wyrazi się wtedy funkcją

$$R(F_i, d_k).$$

Zadaniem naszym jest ten koszt uczynić jak najmniejszym.

Jeśli proces aktywności słońca opisuje się dystrybuantą

$$a_1 F_1 + \dots + a_n F_n,$$

a my posługujemy się funkcją zrandomizowaną

$$\beta_1 d_1 + \dots + \beta_m d_m \quad (3)$$

(β_i oznacza prawdopodobieństwo wybrania funkcji decyzyjnej d_i), to średnio koszt całkowity wyrazi się formą biliniową

² Koszt wykorzystywania pomiarów i stratę wynikającą ze złej decyzji.

$$\sum_{i=1, j=1}^{i=n, j=m} a_i \beta_j R(F_i, d_j),$$

którą dla skrót u oznaczymy

$$R_{ij} a_i \beta_j.$$

Maksimum formy $R_{ij} a_i \beta_j$ ze względu na a_i przy ustalonych β_j nazywamy kosztem gwarantowanym przez funkcję decyzyjną (3) (zrandomizowaną).

Funkcję decyzyjną zrandomizowaną $b_1 d_1 + \dots + b_n d_n$ nazywamy najlepszą jeżeli gwarantuje minimalny koszt

$$\max_{a_i, i=1, \dots, n} R_{ij} a_i \beta_j = \min_{\beta_1, \dots, \beta_m} \max_{a_i, i=1, \dots, n} R_{ij} a_i \beta_j.$$

Aby zilustrować rachunkowe trudności przy efektywnym znajdowaniu współczynników b_j , rozwiążemy prosty przykład przyjmując $i=2, j=2$. Przy większych wartościach i, j trudności wzrastają podobnie jak przy przechodzeniu od dwóch równań liniowych z dwiema niewiadomymi do większej liczby równań i niewiadomych.

Przyjmijmy $R_{ij} a_j \beta_i = a_1 \beta_1 + a_2 \beta_2$.

Mamy wyznaczyć $\min_{\beta_1, \beta_2} \max_{a_1, a_2} a_1 \beta_1 + a_2 \beta_2$.

Ekstremum mieszane *minimum maximorum* wyznaczamy podobnie jak zwyczajne, przez przyrównanie pochodnych do zera. Aby ułatwić sobie analityczne operowanie zmiennymi $a_1, a_2, \beta_1, \beta_2$, które są, jak wiemy dodatnio unormowane³, najlepiej zamiast tych zmiennych użyć zmiennych $\sin^2 \alpha, \cos^2 \alpha, \sin^2 \beta, \cos^2 \beta$.

Różniczkując formę

$$\sin^2 \alpha \cdot \sin^2 \beta + \cos^2 \alpha \cdot \cos^2 \beta, \quad (4)$$

otrzymujemy na ekstrema układ równań

$$\begin{aligned} 2(\sin^2 \beta - \cos^2 \beta) \sin \alpha \cos \alpha &= 0, \\ 2(\sin^2 \alpha - \cos^2 \alpha) \sin \beta \cos \beta &= 0. \end{aligned}$$

Układ ten ma pierwiastki:

- 1) $\sin \alpha = \sin \beta = 0$,
- 2) $\cos \alpha = \cos \beta = 0$,
- 3) $\sin \alpha = \cos \beta = 0$,
- 4) $\cos \beta = \sin \alpha = 0$,
- 5) $\sin^2 \alpha = \cos^2 \alpha = \sin^2 \beta = \cos^2 \beta = \frac{1}{2}$.

Łatwo sprawdzić, że pierwiastkom 1) i 2) odpowiada maksimum, pier-

³ To znaczy

$$\begin{aligned} \alpha_1 + \alpha_2 &= 1, & \alpha_1 &\geq 0, & \alpha_2 &\geq 0, \\ \beta_1 + \beta_2 &= 1, & \beta_1 &\geq 0, & \beta_2 &\geq 0. \end{aligned}$$

wiastkom 3) i 4) — minimum formy 4. Pierwiastkom 5) odpowiada ekstremum mieszane. Znaleźliśmy więc ekstremalne wartości $b_1 = \frac{1}{2}$, $b_2 = \frac{1}{2}$.

Wynik ten można było przewidzieć, gdybyśmy przedtem wyjaśnili, że forma $a_1\beta_1 + a_2\beta_2$ odpowiada zagadnieniu prognozy orla i reszki w przykładzie, o którym wspominaliśmy. Współczynniki b_1, b_2 mają znaczenie prawdopodobieństw, z jakimi powinniśmy przepowiadać orla i reszkę.

Oczywiście teoria decyzyjna została przedstawiona w sposób bardzo pobieżny i szkicowy.

Wspomnijmy jeszcze, że są następujące drogi ominięcia trudności szacowania straty gospodarczej ze złej prognozy. Koszt wykorzystania badań i jakość prognozy można mierzyć w różnych jednostkach i powyższą teorię uprawiać jak gdyby w zakresie liczb zespolonych. Wtedy zamiast optymalnej funkcji decyzyjnej otrzymujemy rodzinę optymalnych nieporównywalnych między sobą funkcji, ostatecznego zaś wyboru dokonujemy w zależności od lokalnych potrzeb i możliwości. Teoria dostarcza jak gdyby ofert, o których wie, że są lepsze od innych możliwości, ale między którymi już nie rozstrzyga.

Inną drogą jest budowa analogicznej, chociaż istotnie różnej teorii opartej, z grubsza mówiąc, na możliwościach finansowych instytucji prowadzącej badania. Chodzi o to, że w praktyce i tak na takie a takie badania przeznaczamy tyle a tyle. Teoria wskazuje natomiast optymalny schemat badań opartych na danym budżecie i, co więcej, pozwala sekwencyjnie badać czy dalsze zwiększenie budżetu jest celowe, czy też budżet ten należy zmniejszać.

Trzeba zaznaczyć, że przy stosowaniu wszystkich wyżej przedstawionych teorii prognoz w praktyce napotykamy dość duże trudności. Wynika to z faktu, że chociaż teorie te są na ogół szeroko i precyzyjnie rozbudowane, to jednak prowadzą często do niezmiernie skomplikowanych rachunków numerycznych lub wymagają trudnych i kosztownych badań.

*Instytut Matematyczny
Polskiej Akademii Nauk*

WYKAZ LITERATURY

1. Dworetzky A., Kiefer J., Wolfowitz I.: *Sequential decision problems for processes with continuous time parameters problems of estimation*. The Ann. of Math. Statistics, T. 24, 1953, p. 403—415.
2. Jagłom A. M.: *Wwiedienije w tieorju stacionarnych sluczajnych funkcji*. Uspiechi Matematicheskich Nauk, T. VII, 5(51), 1952, s. 3 — 168.
3. Moran P. A. P.: *Some experiments and the prediction of sunspot numbers*. CCIR, London, 1953.
4. *Prediction of solar index*. Preliminary on recommendation nr 69, CCIR, London 1953.
5. Wald A.: *Sequential decision functions*, N. Y., 1950, p. 1—176.
6. Wiener N.: *Extrapolation, interpolation and smoothing of stationary time series*. New York, 1949.

К. УРБАНИК, А. ЗЕМБА

ПРОГНОЗ СОЛНЕЧНОЙ АКТИВНОСТИ

Резюме

Изменения электромагнитной солнечной активности, вызывающие в особенности помехи приёма радиопередачи, характеризуются вообще числами Вольфа, описывающими распределение и количество солнечных пятен. Поэтому прогноз солнечной активности сводится к прогнозу чисел Вольфа. Точная зависимость чисел Вольфа от времени — нам неизвестна. В первом приближении принимается, что изменения чисел Вольфа имеют случайный характер. Это позволяет использовать в прогнозе аппарат стохастических процессов. В статье вкратце обсуждены пробабилистические методы прогноза числа Вольфа. Лучшее всего исследован теоретически метод линейного прогноза Колмогорова - Винера, который тем лучше, чем большее число наблюдательных данных он использует. Однако большое число наблюдательных данных ведёт к длинными и кропотливым вычислениям. Не следует поэтому пренебрегать методами прогноза, использующими небольшое число данных и дающими сравнительно хорошие результаты. В статье приведен такой метод, учитывающий стоимость опыта и основанный на теории децизионных функций.

K. URBANIK, A. ZIĘBA

PREDICTION OF SOLAR ACTIVITY

Summary

Electromagnetic changes in solar activity, which cause disturbances in radio reception, are characterized in a general manner by Wolf numbers. These numbers describe the location and number of sunspots. Thus, prediction of solar activity is reduced to the prediction of Wolf numbers. The exact dependency of Wolf numbers on time is not known. As a first approximation, it is assumed that the change in the Wolf numbers is of a chance nature. This permits stochastic process apparatus to be used for predictions. In this paper, probalistic methods for predicting Wolf numbers are briefly discussed. The method most investigated is the Kolmogorov-Wiener linear method of prediction. In this method, the greater the amount of data employed, the better the results obtained. A great quantity of observed data, however, lead to long and timeconsuming calculations. Therefore, prediction methods employing not too great quantitative data, and giving relatively good results, are not without practical importance. In this paper, one such method, based on the theory of decision functions, is described. The cost of the experiment has also been taken into account.

621.392.52.

J. LENKOWSKI

Miara przydatności filtrów pasmowych o charakterystyce Czebyszewa

Rękopis dostarczono 6. 6. 1955

W pracy przedstawiono koncepcję pewnego kryterium przydatności filtrów, związanego z kształtem charakterystyki. Kryterium to jest miarą zakłóceń od sąsiedniego kanału względem zniekształceń liniowych sygnału pożądanego. Można zastosować to kryterium nie tylko do filtrów o charakterystyce Czebyszewa ale i do innych filtrów o charakterystykach maksymalnie płaskiej lub maksymalnie ostrej. Porównując ze sobą współczynniki przydatności, obliczone dla różnych charakterystyk, dochodzimy do wniosku, że najlepszy jest z tego punktu widzenia filtr o charakterystyce Czebyszewa.

1. WSTĘP

Znany jest sprawdzian dobroci wzmacniaczy pasmowych (wzmacniaczo-filtrów) w postaci iloczynu $K \cdot B$, gdzie K jest wzmocnieniem w środku pasma, B zaś szerokością pasma przepuszczania.

Wykażemy niżej, że można pomyśleć jeszcze jeden sprawdzian, przedstawiający miarę zakłóceń pochodzących od sąsiedniego kanału w stosunku do zniekształceń liniowych własnego sygnału przez dany filtr. Sprawdzian ten związany jest więc raczej z pojęciem odległości między sygnałem oryginalnym a zniekształconym.

Aby dojść do zamierzonego celu, należy się jednak cofnąć do zagadnienia aproksymacji w syntezie filtrów. Jednym z etapów syntezy jest aproksymacja; chodzi tu o to, że z tych lub innych rozważań wynika pewien pożądaný przebieg funkcji przenoszenia filtru, będącej na przykład stosunkiem napięcia na wyjściu do napięcia na wejściu. W przypadku sygnału o modulacji amplitudy obwiednią akustyczną, przebieg fazy funkcji przenoszenia ma znaczenie drugorzędne; można tu więc podać jedynie pożądaný przebieg modułu funkcji przenoszenia. Nie wchodząc w słuszność sprawy, zazwyczaj jako taki pożądaný kształt podaje się przebieg prostokątny (rys. 1) lub jego odwrotność (rys. 2). Otóż konieczność aproksymacji wynika stąd, że taki kształt przebiegu modułu funkcji

przenoszenia nie daje się zrealizować w jakimkolwiek fizycznie możliwym układzie *RLC*.

Układy realne mają w przypadku ogólnym funkcje przenoszenia będące wymiernymi funkcjami parametru częstotliwości zespolonej

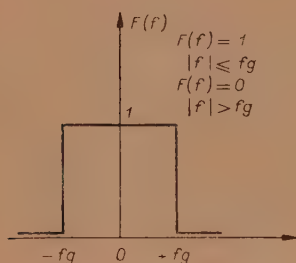
$$p = \sigma + j\omega,$$

gdzie

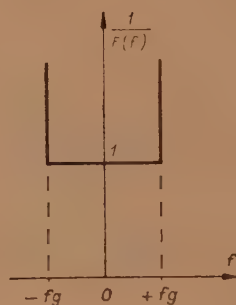
σ — jest tłumieniem,

ω — częstotliwością.

Dla układów wzmacniaczo-filtrów będących kaskadowym połączeniem lamp (lub tranzystorów) i filtrów o strukturze obwodów sprzężonych

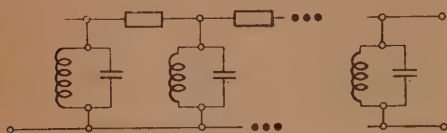


Rys. 1. Żądana funkcja przenoszenia.



Rys. 2. Odwrotność funkcji przenoszenia.

(rys. 3), odwrotność funkcji przenoszenia $1/F(x)$ ma jeszcze prostszy kształt, bo jest funkcją wymierną, całkowitą, lub po prostu wielomianem przebiegu z rys. 1 lub z rys. 2, musimy go aproksymować.



Rys. 3. Obwody sprzężone

Sprawa przybliżonego przedstawienia dowolnych funkcji za pomocą prostszych funkcji (a przede wszystkim wielomianów) stanowi treść pewnej gałęzi analizy matematycznej [1—3].

Istotną rolę odgrywa tu ustalenie zbioru funkcji, w którym aproksymacja ma być dokonana, a do którego należeć powinny prostokątna zadana funkcja przenoszenia, jak również i wielomiany algebraiczne. Zbiór ten tworzy się na podstawie pewnego kryterium wspólnego dla wszystkich jego elementów. Drugą, również ważną sprawą jest wyjaśnienie sensu twierdzenia, że pewna funkcja, będąca elementem omawianego zbioru, jest bliska innej, należącej do tegoż zbioru. Inaczej mówiąc, niezbędne jest tu wprowadzenie definicji odległości. Zbiór funkcji zadaną odległością tworzy przestrzeń metryczną, a same elementy są punktami tej przestrzeni.

Dla naszych zamierzeń odpowiednia jest tak zwana przestrzeń C , która jest zbiorem wszystkich funkcji ciągłych, zadanych na odcinku $[a, b]$. Odległość od siebie dwóch elementów przestrzeni x, y definiujemy tu jako:

$$D(x, y) = \max_{a \leq t \leq b} |x(t) - y(t)|.$$

Do powyższych rozważań należy jeszcze dodać, że przestrzeń C jest również przestrzenią liniową. Znaczy to, że istnieją tu operacje dodawania poszczególnych elementów i mnożenia ich przez liczby według zwykłych zasad.

Poza tym przestrzeń liniowa staje się normowana z chwilą zdefiniowania normy, określającej w pewien sposób wielkości każdego jej elementu. Przestrzeń C posiada normę swego elementu x określoną następująco:

$$\|x\| = \max_{a \leq t \leq b} |x(t)|.$$

Łatwo jest z tej zależności zauważyć, że poprzednio wprowadzona odległość wiąże się z pojęciem normy w sposób następujący:

$$D(x, y) = \|x - y\| = \max_{a \leq t \leq b} |x(t) - y(t)|. \quad (1)$$

W omówionej powyżej liniowej normowanej przestrzeni C można obecnie sformułować najogólniej zagadnienie aproksymacji tak, jak to zrobiono poniżej.

Dany jest element x tej przestrzeni, należy zaś znaleźć liczby $\lambda_1, \lambda_2, \dots, \lambda_n$ takie, aby odległość

$$D(x, \lambda_1 g_1 + \lambda_2 g_2 + \dots + \lambda_n g_n) = \|x - \lambda_1 g_1 - \dots - \lambda_n g_n\|$$

była możliwie najmniejsza, jeżeli $g_1, g_2, \dots, g_n$ jest zespołem n liniowo niezależnych¹ elementów tejże samej przestrzeni. Istnieje dowód, że tak sformułowane zagadnienie aproksymacji posiada rozwiązanie, to znaczy istnieją liczby $\lambda_1, \lambda_2, \dots, \lambda_n$ ([1], s. 16).

Rozwiązanie to, ogólnie biorąc, nie jest nawet jednoznaczne. Jest ono jednoznaczne w przypadku, gdy elementy $g_1, g_2, \dots, g_n$ mają postać $t^0, t^1, t^2, \dots, t^n$, to znaczy, gdy element $x(t)$ aproksymujemy za pomocą wielomianu algebraicznego. Wielomian najbliższy w sensie (1) danej $x(t)$ będziemy nazywali wielomianem najlepszego przybliżenia. Po raz pierwszy takie wielomiany były rozpatrzone przez rosyjskiego matematyka P. Ł. Czebyszewa.

¹ Liniową niezależność elementów $g_1, g_2, \dots, g_n$ rozumiemy w ten sposób, że jeśli istnieje zależność

$$\alpha_1 g_1 + \alpha_2 g_2 + \dots + \alpha_n g_n = 0,$$

gdzie $\alpha_1, \alpha_2, \dots, \alpha_n$ są liczbami, to wynika z niej

$$\alpha_1 = \alpha_2 = \dots = \alpha_n = 0$$

2. APROKSYMACJA KRZYWEJ PROSTOKĄTNEJ

Wracając do postawionego we wstępie problemu aproksymacji wielomianem algebraicznym krzywej prostokątnej (rys. 2), napotykamy zasadniczą trudność, że nie należy ona do przestrzeni C . W zasadzie istnieje tylko jedna droga do pominięcia tej trudności; polega ona na zmianie kształtu żądanej funkcji przenoszenia na inny, będący funkcją ciągłą na danym odcinku $[a, b]$.

Idąc za klasycznym rozwiązaniem możemy tu ograniczyć aproksymację do odcinka $[-fg, fg]$, stanowiącego pasmo przepuszczania, przez co uzyskuje się pożądaną funkcję spełniającą warunek ciągłości. W tym przypadku problem formułuje się zazwyczaj następująco.

Wśród wszystkich wielomianów n -tego stopnia ze współczynnikiem p_n przy x^n znaleźć taki, dla którego

$$D = \max_{(-1 \leq x \leq 1)} |Q_n^*(x)|$$

jest najmniejsze.² Biorąc pod uwagę jeszcze fakt, że dla $|x| \rightarrow \infty$, $Q_n^*(x) \rightarrow \infty$ będziemy rozumieli, że w rezultacie tak postawionego zagadnienia, aproksymować będziemy krzywą prostokątną przedstawioną na rys. 2.

Aby móc jednak zastosować ogólne twierdzenie o wielomianach najlepszego przybliżenia, składnik n -tej potęgi wielomianu $Q_n^*(x)$

$$p_n x^n$$

przyjmijmy jako funkcję aproksymowaną, a wielomianem aproksymacyjnym będą wówczas pozostałe składniki $Q_n^*(x)$

$$P_{n-1}(x) = -[Q_n^*(x) - p_n x^n] = -(p_{n-1} x^{n-1} + p_{n-2} x^{n-2} + \dots + p_0).$$

Wielomian aproksymacyjny jest więc $(n-1)$ -stopnia.

Warunkiem koniecznym i dostatecznym, aby wielomian $P_{n-1}(x)$ był wielomianem najlepszego przybliżenia, jest istnienie tak zwanego alternansu Czebyszewa [1—3], składającego się z co najmniej $(m+2)$ punktów odcinka a, b , na którym dokonywana jest aproksymacja

$$a \leq x_1 < x_2, \dots, < x_{m+2} \leq b,$$

gdzie

m — stopień wielomianu aproksymacyjnego
(w naszym przypadku $m = n - 1$).

W tych punktach różnica funkcji aproksymowanej i wielomianu aproksymalnego osiąga wartość D ze znakiem $(+)$ i $(-)$ na przemian. Jak

² Jest tu mało istotną sprawą przyjęcie odcinka $[-1, 1]$ zamiast $[-fg, fg]$. Również łatwo da się aproksymować wartość 1 (jak na rys. 1) zamiast zera, co uzyskamy przez dodanie jedności do $Q_n^*(x)$.

wiadomo [2, 3] wielomianem, który zadość czyni tym warunkom, jest w naszym przypadku

$$Q_n^*(x) = \frac{p_n}{2^{n-1}} \cos(n \arccos x) = \frac{p_n}{2^{n-1}} T_n(x), \quad (2)$$

gdzie $T_n(x)$ jest wielomianem Czebyszewa pierwszego rodzaju, n -tego stopnia. Odległość

$$D = \frac{p_n}{2^{n-1}}. \quad (3)$$

Na marginesie rozwiązanego w ten sposób zagadnienia należy jednak zauważyć, że opuszczony tu został może najistotniejszy szczegół związany z samym celem stosowania filtrów, mianowicie zagadnienie selektywności, zagadnienie kształtu zboczy. Wynika to oczywiście z tego, że zbocza nie zostały objęte aproksymacją i możemy tu jedynie stwierdzić, że $Q_n^*(x)$, posiadając w nieskończoności biegun n -tego stopnia, dla dostatecznie dużych x zachowuje się w przybliżeniu jak jego składnik n -tej potęgi $p_n x^n$.

3. APROKSYMACJA PROSTEJ Z WARUNKIEM POBOCZNYM OSIĄGNIĘCIA ZAŁOŻONEJ WARTOŚCI W PUNKCIE NA ZEWNĄTRZ PRZEDZIAŁU

Istnieje jeszcze inny sposób pominięcia trudności aproksymacji krzywej kształtu z rys. 2 w przestrzeni C . Mianowicie aproksymację prostokąta można zastąpić następującym zagadnieniem [2]. Wśród wszystkich wielomianów n -tego stopnia $Q_n(x)$ przyjmujących w punkcie ξ (dla $|\xi| > 1$) wartość z góry zadaną η , znaleźć taki, który najmniej odchyła się od zera na odcinku $-1 < x < 1$. Wielomian ten musi mieć kształt

$$Q_n(x) = \eta + (x - \xi) R(x),$$

gdzie $R(x)$ jest wielomianem $(n-1)$ stopnia.

Poszukiwany wielomian powinien zapewnić minimalną wartość wyrażenia

$$D_q = \max_{-1 < x < 1} \eta - (x - \xi) R(x) = \max_{-1 < x < 1} x - \xi; \quad \left| \frac{\eta}{x - \xi} - R(x) \right|. \quad (4)$$

Nasuujące się zadanie można wobec powyższego rozumieć w sposób następujący: należy znaleźć najlepsze przybliżenie funkcji

$$f(x) = \frac{\eta}{x - \xi}$$

za pomocą wielomianu $[-R(x)]$ stopnia $(n-1)$. W porównaniu z zagadnieniem rozpatrzonym w rozdziale 2 istnieje tu pewna różnica. Polega ona

na odmiennej definicji odległości D_q od stosowanej w przestrzeni C . Chodzi tu oczywiście o czynnik

$$q(x) = |x - \xi|,$$

który nazwiemy funkcją wagi.³

Odległość D_q określona przez zależność (4) jest w związku z tym odległością ważoną.

Okazuje się, że przyjęcie ważonej odległości nie wprowadza większych zmian w samym zagadnieniu. Mianowicie, poza dającym się udowodnić faktem istnienia jedynego wielomianu najlepszego przybliżenia ([2], s. 223), koniecznym i wystarczającym warunkiem do tego, aby dany wielomian był właśnie taki, jest istnienie alternansu Czebyszewa, analogicznie jak poprzednio. Konkretnie, ważona różnica

$$q(x) [f(x) + R(x)]$$

musi na odcinku $[-1, 1]$ osiągnąć wartość D_q w co najmniej $(n+1)$ punktach, kolejno zmieniając znak.

Porównując ten wniosek z wynikiem rozważań w rozdziale poprzednim dochodzimy do wniosku, że wielomiany $Q_n(x)$ i $Q_n^*(x)$ (2) różnić się mogą najwyżej stałym czynnikiem.

Możemy łatwo ustalić wartość tego współczynnika, korzystając z warunku

$$Q_n(\xi) = \eta.$$

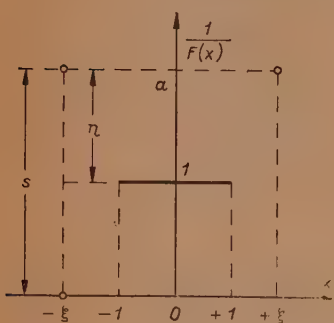
Warunek ten, łącznie z poprzednio omówionym będzie spełniony, gdy $Q_n(x)$ będzie miało kształt

$$Q_n(x) = \frac{\eta}{|T_n(\xi)|} T_n(x). \quad (5)$$

Maksymalne odchylenie na odcinku $-1 \leq x \leq 1$ jest

$$D_q = \frac{\eta}{|T_n(\xi)|}. \quad (6)$$

Rys. 4 Funkcja przenoszenia z określonym jednym punktem na zboczu.



Pozostaje obecnie dokonać przesunięcia równoległego o jedność wzdłuż osi rzędnych, aby otrzymać wielomian, odpowiadający wymaganiom (rys. 4).

³ Funkcja wagi $q(x)$ z założenia powinna odpowiadać pewnym warunkom, a mianowicie na odcinku $[-1, 1]$ jest funkcją ciągłą dodatnią, spełniającą nierówność $0 < m \leq q(x) \leq M$, gdzie m i M — pewne dodatnie liczby.

Biorąc pod uwagę, że poszukiwany wielomian musi być parzysty, otrzymamy ostatecznie

$$G_{2m}(x) = 1 + \frac{s-1}{|T_{2m}(\xi)|} T_{2m}(x) \quad (7)$$

oraz

$$D_q = \frac{s-1}{|T_{2m}(\xi)|} \quad (8)$$

4. MIARA DOBROCI FILTRU Z CHARAKTERYSTYKĄ CZEBYSZEWA

Aproksymacja przeprowadzona w rozdziale 3 w przeciwstawieniu do rozwiązania klasycznego (rozd. 2) uwzględnia przebieg zboczy krzywej modułu odwrotności funkcji przenoszenia. Uwidoczniony tu zostaje związek istniejący między zniekształceniem sygnału użytecznego, przepuszczanego przez filtr, a zakłóceniem sygnału użytecznego przechodzącym na przykład od sygnału znajdującego się w kanale sąsiednim.

Rozważmy wyrażenie (8)

$$D_q = \frac{s-1}{|T_{2m}(\xi)|},$$

które przepiszemy nieco inaczej jako

$$\frac{s-1}{D_q} = |T_{2m}(\xi)|. \quad (9)$$

Licznik lewej strony (9) możemy rozumieć jako stosunek: kwadrat sygnału użytecznego/zakłócenie od sygnału w kanale sąsiednim ($x=\xi$).⁴ Sama odchyłka D_q może być znów uważana za kwadrat względnego zniekształcenia liniowego sygnału użytecznego w pasmie przepuszczania.⁵

Wówczas z wyrażenia (9) stwierdzamy, że kwadrat stosunku: sygnał do zakłócenia do względnych zniekształceń liniowych sygnału jest określony jedynie przez stopień złożoności filtru (indeks m) oraz położenie (ξ) zakłócenia względem środka pasma przepuszczania. Zależność tę można jeszcze bardziej uprościć pisząc w postaci

$$\ln |T_{2m}(\xi)| = \ln(s-1) - \ln D_q.$$

Wprowadzając zaś oznaczenia dla tłumienia względnego zakłóceń

$$\ln(s-1) = S \quad (10)$$

oraz względnego liniowego zniekształcenia sygnału

$$\ln D_q = L, \quad (11)$$

⁴ Dla $s \gg 1$ (co w praktyce jest spełnione) $s-1 \approx s$.

⁵ Dla $D_q \ll 1$.

otrzymamy ostatecznie

$$S - L = \ln |T_{2m}(\xi)|. \quad (12)$$

Wielkość określoną równaniem (9) bądź (12) możemy przyjąć jako miarę przydatności omawianego filtru pasmowego. Miara ta ustala różnicę tłumienia względnego zakłócenia i względnych zniekształceń liniowych sygnału. Mamy tu więc powiązane ze sobą dwa istotne elementy przy przechodzeniu sygnału przez filtr. Fakt powiązania zniekształceń własnych i zewnętrznych ze strukturą filtru pozwala rozważyć nawet w pewnym uproszczonym sensie zagadnienie jego optymalizacji. Dla dokonania tego można przyjąć, że sygnał użyteczny ma widmo amplitudowe o przebiegu prostokątnym zawartym w pasmie $-1 \leq x \leq 1$. Jako kryterium optymalizacji ustalimy, że odległość sygnału na wyjściu z filtru względem sygnału wejściowego, ma być możliwie najmniejsza. Logiczne jest tu przyjęcie odległości w sensie metryki przestrzeni C . W tym przypadku jednak dostajemy trywialny rezultat, że

$$D_q = \frac{s-1}{|T_{2m}(\xi)|}$$

ma osiągnąć możliwie najmniejszą wartość, co znaczy, że szczyt krzywej przepuszczania ma być możliwie płaski. Zniekształcenie zewnętrzne jest tu całkowicie pominięte. Aby je uwzględnić, problem optymalnego filtru sformułujemy ostatecznie tak: poszukujemy D_q możliwie najmniejszego, przy pobocznym warunku $s = \text{const}$. Prowadzi to w sposób oczywisty do wniosku, że $|T_{2m}(\xi)|$ musi być możliwie duże, a zatem albo ξ możliwie małe, albo m możliwie duże.

5. PRZYKŁADY ZASTOSOWANIA OMÓWIONEGO KRYTERIUM DO FILTRÓW O INNYCH CHARAKTERYSTYKACH

Zastosujemy obecnie kryterium (9) do innych charakterystyk filtrów, a to celem porównania ich z poprzednio omówionymi.

Najpierw rozważmy tak zwaną charakterystykę maksymalnie płaską, którą w skrócie oznaczmy literą B . Wprowadzimy tu pojęcie odchylenia D w pasmie przepuszczania $[-1, 1]$ jak na rys. 5. Wielomian o maksymalnie płaskim przebiegu, odpowiadający załączonemu rysunkowi, ma więc postać

$$G_{2m}(x) = 1 - D + 2Dx^{2m}. \quad (13)$$

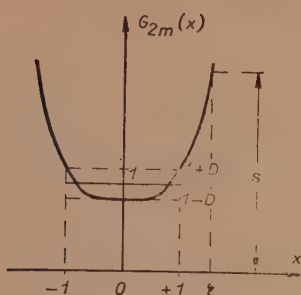
Dla $x = \xi$, gdzie $|\xi| > 1$

$$G_{2m}(\xi) = 1 - D + 2D\xi^{2m} = s,$$

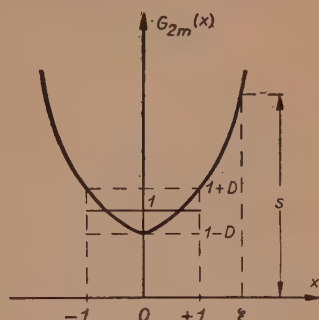
skąd

$$\frac{s-1}{D} = 2\xi^{2m} - 1. \quad (14)$$

Jako drugi przykład weźmiemy tak zwaną charakterystykę maksymalnie ostrą, którą oznaczmy literą A. Odchylenie D da się wprowadzić



Rys. 5. Odchylenie w pasmie przepuszczania charakterystyki B.



Rys. 6. Odchylenie w pasmie przepuszczania charakterystyki A.

analogicznie jak poprzednio, co jest uwidocznione na rys. 6. Odpowiedni wielomian ma kształt

$$G_{2m}(x) = 1 - D + 2D \frac{\sinh^2(m \operatorname{ar} \sinh x)}{\sinh^2(m \operatorname{ar} \sinh 1)}. \quad (15)$$

Dla $x = \xi$ mamy

$$G_{2m}(\xi) = s = 1 - D + 2D \frac{\sinh^2(m \operatorname{ar} \sinh \xi)}{\sinh^2(m \operatorname{ar} \sinh 1)},$$

skąd

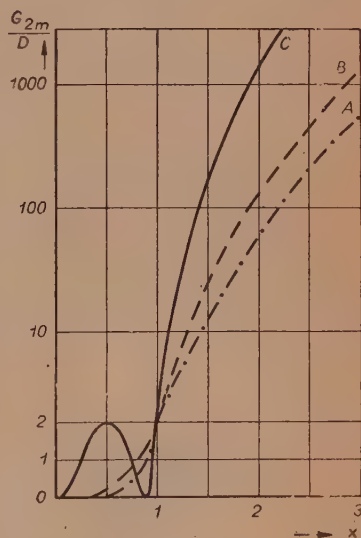
$$\frac{s-1}{D} = 2 \frac{\sinh^2(m \operatorname{ar} \sinh \xi)}{\sinh^2(m \operatorname{ar} \sinh 1)} - 1. \quad (16)$$

Porównując ze sobą wyrażenia (9), (14) i (16) dochodzimy do wniosku, że

$$\left[\frac{s-1}{D} \right]_A < \left[\frac{s-1}{D} \right]_B < \left[\frac{s-1}{D} \right]_C,$$

gdzie (C) oznacza charakterystykę Czebyszewa. Nierówności powyższe zilustrowane na

rys. 7, gdzie przedstawiony jest przebieg $G_{2m}(x)$ dla charakterystyk A, B, C, dla $m=3$, przy czym oś rzędnych wyskalowana jest w wielkości D



Rys. 7. Znormalizowana krzywa $G_6(x)$.

WYKAZ LITERATURY

1. Achiezer N. I.: *Lekcii po teorii aproksimacji*, Moskwa 1947.
2. Gonczarow W. T.: *Teoriya interpolirovaniia i pribliżeniia funkcij*. Moskwa 1954.
3. Natanson I. P.: *Konstruktivnaja teoriya funkcij*. Moskwa 1949.

И. ЛЕНКОВСКИ

МЕРА ДОВРОТНОСТИ ПОЛОСОВЫХ ФИЛЬТРОВ
С ХАРАКТЕРИСТИКОЙ ЧЕБЫШЕВА

Резюме

Из различных рассуждений вытекает некоторая желательная форма передаточной функции, являющейся соотношением выходного напряжения ко входному. В случае сигнала с модуляцией амплитуды акустической частотой достаточно, вообще говоря, определение желательной формы характеристики самого модуля функции. Обыкновенно форма эта оказывается такой, какой она представлена на рис. 1. Ввиду невозможности его осуществления при помощи конечного числа элементов RLC возникает проблема аппроксимации. Для систем фильтро-усилительных, являющихся каскадным соединением ламп (или транзисторов) и фильтров со структурой связанных контуров (рис. 3), обратная величина передаточной функции является алгебраическим многочленом, при помощи которого можно аппроксимировать желательную форму характеристики (рис. 2).

Исполняя известным образом аппроксимацию в пространстве C непрерывных функций, мы принуждены ограничить отрезок аппроксимации к $[-f_0, f_0]$ т.е. к самой полосе пропускания фильтра. Иначе говоря, мы совершаем здесь аппроксимацию собственно не прямоугольной кривой, изображенной на рис. 2, но только её плоской вершины.

Отсюда видно, что здесь пропущено существенное обстоятельство, связанное с самой целью применения фильтров, а именно избирательность на боковых частях амплитудной характеристики.

Возможно однако провести аппроксимацию отрезка прямой линии в пространстве C в полосе пропускания с одновременным возложением условия избирательности вне отрезка аппроксимации, как это представлено на рис. 4. В результате мы получаем желаемый аппроксимационный многочлен вида

$$G_{2m}(x) = 1 + \frac{s-1}{|T_{2m}(\xi)|} T_{2m}(x) \quad (1)$$

где

 x — относительная частота, s — значение достигаемое многочленом в данной точке, лежащей вне полосы пропускания, $T_{2m}(x) = \cos(2m \arccos x)$ — многочлен Чебышева $2m$ -той степени, первого рода.Максимальное отклонение многочлена $G_{2m}(x)$ от аппроксимированной прямой в полосе пропускания выражается

$$D_q = \frac{s-1}{|T_{2m}(\xi)|}.$$

Если представить это выражение следующим образом

$$\frac{s-1}{D_q} = |T_{2m}(\xi)| \quad (2)$$

то полученную величину можно принять как известного рода меру пригодности фильтров с характеристикой Чебышева.

Числитель левой стороны выражения (2) мы можем понимать (для $s \gg 1$, что на практике действительно имеет место) как отношение:

$$\left[\frac{\text{полезный сигнал}}{\text{помехи от сигнала в соседнем канале } (x=\xi)} \right]^2.$$

Знаменатель же левой стороны выражения (2), (D_q) можно считать квадратом относительных линейных помех от полезного сигнала в полосе пропускания (для $D_q \ll 1$).

Из выражения (2) мы можем констатировать, что отношение

$$[\text{сигнал/помехи}] : [\text{относительное линейное искажение сигнала}]$$

определяется единственно степенью сложности фильтра (индекс x), а также положением (ξ) помех относительно середины полосы пропускания. Мы имеем таким образом зависимость двух существенных элементов передачи сигнала фильтром: внешнего и собственного искажения.

Критерий (2) можно применить к фильтрам с другими (кроме Чебышева) характеристиками. Так напр. для фильтра с наиболее плоской характеристикой (B) рис. 5 — мы получаем

$$\left[\frac{s-1}{D} \right]_B = 2\xi^{2m} - 1 \quad (3)$$

а для фильтра с наиболее острой характеристикой — рис. 6.

$$\left[\frac{s-1}{D} \right]_A = 2 \frac{\sinh^2 (m \operatorname{arc} \sinh \xi)}{\sinh^2 (m \operatorname{arc} \sinh 1)} - 1 \quad (4)$$

Сравнивая выражения (2), (3) и (4), мы получаем следующее неравенство

$$\left[\frac{s-1}{D} \right]_A < \left[\frac{s-1}{D} \right]_B < \left[\frac{s-1}{D} \right]_C$$

где

C — обозначает характеристику Чебышева.

Неравенства эти пояснены на рис. 7, на котором представлены графики аппроксимационных многочленов $G_{2m}(x)$ для характеристик A, B, C для $m=3$ причём ось ординат в масштабе величины D.

J. LENKOWSKI

FIGURE OF MERIT FOR BAND-PASS FILTERS WITH A TCHEBYSCHEFF CHARACTERISTIC

Summary

For various reasons, it is desirable that the transfer function has a certain shape. In the case of an audio amplitude modulated signal, it is, in general, sufficient to take as a desirable waveform the one shown in Fig. 1. Since it is impossible to obtain this waveform with a finite number of RLC elements, the question becomes an approximation problem. For amplifier-filter systems in which the valves (or transistors) and composite filters are coupled in cascade (Fig. 3), the inverse of the transfer function is an algebraic polynomial. With the help of this polynomial, the desired waveform of Fig. 2 may be approximated.

By applying the usual approximation method for the continuous functions in space C , we are forced to limit the approximation interval to $[-f_g, f_g]$, i. e. to the band-pass of the filter. In other words, we are not actually approximating the rectangular curve shown in Fig. 2, but only its flat peak.

It is thus seen that here have been neglected the most essential details connected with the very reason for using the filters, namely, the selectivity at the sides of the band-pass curve.

The straight line segment in the band-pass in space C can, however, be approximated whilst, at the same time, the selectivity conditions outside the approximation interval, as shown in Fig. 4 are imposed. As the result, we obtain the desired approximation polynomial in the form

$$G_{2m}(x) = 1 + \frac{s-1}{|T_{2m}(\xi)|} T_{2m}(x) \quad (1)$$

where

x — relative frequency,

s — the value of the polynomial at the given point lying outside the pass band,

$T_{2m}(x) = \cos(2m \arccos x)$ — Tchebycheff polynomial of the $2m$ degree and the first kind.

The maximum deviation of the polynomial $G_{2m}(x)$ from the approximated straight line in the band-pass is

$$D_q = \frac{s-1}{|T_{2m}(\xi)|}$$

If we write the last expression in the form

$$\frac{s-1}{D_q} = |T_{2m}(\xi)| \quad (2)$$

we can treat the quantity obtained as a kind of figure of merit for filters with a Tchebycheff characteristic. Thus, the numerator of the left-hand side of (2) can be regarded as the ratio (for $s \gg 1$, which is the case in practice):

$$\left[\frac{\text{effective signal}}{\text{interference caused by signal in neighbouring channel } (x=\xi)} \right]^2$$

The denominator of the left-hand side of (2), D_q can, in turn, be regarded as the square of relative linear distortion of the effective signal in the pass band

(for $D_q \ll 1$).

From expression (2), it can be seen that the value of the ratio

[signal/interference] : [relative linear distortion of the signal]

is determined only by the degree of complexity of the filter (index m) and the position (ξ) of the interference with respect to the centre of the band-pass. Thus the two essential elements of the transmission of the signal by the filter are connected with each other: the external distortion and self distortion.

Criterion (2) can be used for filters having a characteristic other than the Tchebycheff one. For example we obtain for a filter with a maximally flat characteristic B in Fig. 5

$$\left[\frac{s-1}{D} \right]_B = 2\xi^{2m} - 1 \quad (3)$$

whilst for a filter whose characteristic has a sharp maximum as in Fig. 6

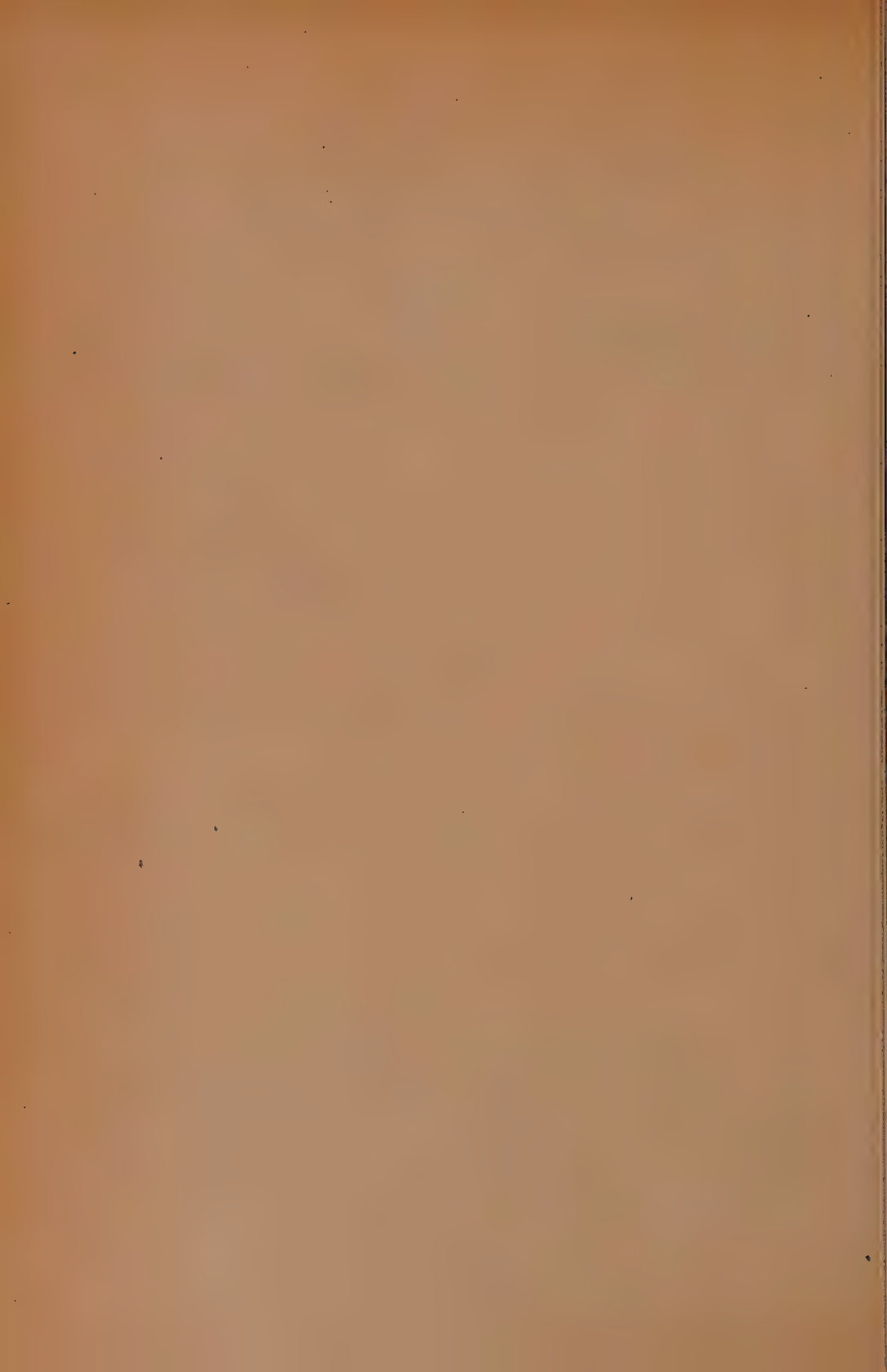
$$\left[\frac{s-1}{D} \right]_A = 2 \frac{\sinh^2 (m \operatorname{arc} \sinh \xi)}{\sinh^2 (m \operatorname{arc} \sinh 1)} - 1 \quad (4)$$

Comparing (2), (3) and (4), we obtain the following inequality

$$\left[\frac{s-1}{D} \right]_A < \left[\frac{s-1}{D} \right]_B < \left[\frac{s-1}{D} \right]_C$$

where C represents the Tchebycheff characteristic.

The above inequalities are illustrated in Fig. 7 in which are shown the curves of the approximation polynomials $G_m(x)$ for characteristics A , B and C for $m=3$; the ordinate axis is drawn in units of the quantity D .



621.3.094.3:621.396.82

R. KULIKOWSKI, A. RYBARSKI

Kryterium zniekształceń kwadratowych przy ograniczonej mocy zakłóceń

Rękopis dostarczono 31. 10. 1955

Wyznaczono optymalne charakterystyki układu liniowego zapewniające przeniesienie sygnału o zadanym kształcie z minimalnymi zniekształceniami kwadratowymi przy założeniu, że średnia moc zakłóceń jest ograniczona. Wyznaczono sygnały pożyteczne i zakłócające, które są najbardziej niekorzystne z punktu widzenia pracy układu. Rozpatrzono również sygnały modulowane i charakterystyki układów selektywnych.

1. WSTĘP

W teorii odbioru sygnałów elektrycznych dużo uwagi poświęca się tzw. filtrom liniowym, których zadaniem jest wydzielanie sygnału pożytecznego spośród zakłóceń. Dla określenia i porównania zdolności filtrujących filtrów odbiorczych liniowych stosuje się różnorodne kryteria. Wybór kryterium uzależnia się zwykle od konkretnych warunków eksploatacyjnych, w jakich pracuje urządzenie odbiorcze.

Jednym z najbardziej rozpowszechnionych kryteriów jest stosunek maksymalnej mocy sygnału do mocy zakłóceń (lub stosunek amplitudy sygnału do skutecznej amplitudy zakłóceń) [1].

Oznaczając przez $S(i\omega)$ widmo sygnału, a przez $Z(i\omega)$ widmo zakłóceń, kryterium powyższe możemy wyrazić jako

$$2\pi \left(\frac{S}{Z} \right)^2 = \frac{\int_{-\infty}^{+\infty} |S(i\omega)K(i\omega)e^{i\omega T}d\omega|^2}{\int_{-\infty}^{+\infty} |Z(i\omega)K(i\omega)|^2 d\omega}, \quad (1)$$

gdzie $K(i\omega)$ jest funkcją przenoszenia układu, T zaś symbolizuje moment czasu, dla którego amplituda sygnału osiąga wartość maksymalną.

Korzystając z nierówności Cauchy-Buniakowskiego można wykazać, że stosunek powyższy osiąga maksymalną wartość dla filtru o funkcji przenoszenia

$$K(i\omega) = \frac{S^*(i\omega) e^{-i\omega T}}{|Z(i\omega)|^2}, \quad (2)$$

gdzie $S(i\omega)$ oznacza funkcję zespolono-sprzężoną z $S(i\omega)$.

Kryterium powyższe okazuje się szczególnie wygodne w przypadku, gdy sygnał S składa się z szeregu identycznych impulsów przedzielonych znacznymi (w porównaniu z szerokością impulsu) odcinkami czasu (jak to ma miejsce na przykład w radiolokacji). W przypadku tym można rzeczywisty sygnał zastąpić pojedynczym impulsem, czyli inaczej mówiąc, okres powtarzania impulsów rozciąga się do nieskończoności.

Należy podkreślić, że kryterium stosunku sygnału do zakłóceń nadaje się szczególnie do układów, w których zależy głównie na wykryciu sygnałów (bądź momentów ich pojawienia się), natomiast nie zależy na dokładnym odtworzeniu kształtów sygnału. W tych urządzeniach, gdzie ważne jest możliwie dokładne odtworzenie kształtu sygnału, szerokie rozpowszechnienie uzyskało kryterium oparte na tzw. błędzie średniokwadratowym [5]:

$$\bar{I} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} [s(t) - u(t)]^2 dt, \quad (3)$$

gdzie

$$u(t) = \int_0^{\infty} e(t-\tau) k(\tau) d\tau,$$

$e(t) = s(t) + z(t)$ — jest sygnałem na wejściu układu,

$s(t)$ — sygnałem pożytecznym,

$z(t)$ — sygnałem zakłócenia,

$k(t)$ — funkcją przejściową, która jest reakcją układu na sygnał impulsowy $\delta(t)$.

Jeśli założymy, że sygnał i zakłócenia opisują się funkcjami stochastycznymi stacjonarnymi (w szerokim sensie), to okazuje się, że wyznaczenie optymalnej funkcji przejściowej $k(t)$ sprowadza się do rozwiązania równania całkowego typu Wienera-Hopfa

$$R_{es}(t) - \int_0^{\infty} k(\tau) R_e(t-\tau) d\tau = 0, \quad (4)$$

gdzie

$$R_{es} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} s(t-\tau) [s(\tau) + z(\tau)] d\tau, \quad (5)$$

$$R_e = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} e(t-\tau) e(\tau) d\tau$$

są tak zwanymi funkcjami korelacyjnymi sygnału $s(t)$ i $z(t)$.

Wynika stąd, że w zagadnieniu filtracji, opartym na kryterium (3), wystarczy znać funkcje korelacyjne sygnału i zakłóceń. Z geometrycznego punktu widzenia zagadnienie sprowadza się do przybliżenia funkcji $s(t)$ i $u(t)$ w Hilbertowskiej przestrzeni funkcyjnej, której własności określa iloczyn skalarny wektorów $s(t)$ i $u(t)$, iloczyn ten pokrywa się natomiast z funkcją korelacyjną.

W kryteriach (1) i (3) operuje się bądź stosunkiem mocy, bądź zniekształceniami sygnału. Kryteria te nie uwzględniają więc wszystkich czynników interesujących w praktyce. Na przykład nie uwzględniają one wzajemnego stosunku między mocą (lub energią) sygnału i zakłóceń na wyjściu układu odbiorczego, a zniekształceniami sygnału.

Dlatego też wydaje się celowe zbadanie kryteriów opartych z jednej strony na zniekształceniu sygnału, z drugiej — operujących mocą (lub energią) sygnału oraz zakłóceń.

W niniejszej pracy ograniczamy się do rozpatrzenia kryteriów, w których dla oceny zniekształceń wykorzystuje się błąd kwadratowy przy ograniczonej mocy (lub energii) zakłóceń na wyjściu układu.

2. KRYTERIUM ZNIEKSZTAŁCEŃ KWADRATOWYCH PRZY OGRANICZONEJ MOCY ZAKŁÓCEŃ

Gdy sygnał $e(t)$ działa na układ liniowy o funkcji przejścia $h(t)$ wyrażającej reakcję układu na skok jednostkowy

$$1(t) = \begin{cases} 1 & \text{dla } t > 0, \\ 0 & \text{dla } t < 0, \end{cases}$$

wówczas na wyjściu układu ukaże się sygnał

$$u(t) = \frac{d}{dt} \int_0^{\infty} e(t-\tau) h(\tau) d\tau.$$

W rozmaitych zagadnieniach wymaga się, aby sygnał wyjściowy z minimalnym zniekształceniem odtwarzał przebieg $x(t)$. Na przykład przy projektowaniu układów przekazujących i wzmacniających mamy zależność

$$x(t) = \text{const} \cdot e(t - t_0)$$

gdzie t_0 jest zadany czas opóźnienia sygnału w układzie. Jako kryterium służące do oceny zniekształceń rozważymy wyrażenie

$$I = \int_0^{\infty} [x(t) - u(t)]^2 dt. \quad (6)$$

Oprócz sygnałów pożytecznych na układ działają często sygnały zakłócające, których efekt działania powinien być dostatecznie *wygaszony*.

Efekt działania zakłóceń można oceniać różnie. Tu ograniczymy się do rozpatrzenia średniej mocy szumów oraz energii zakłóceń na wyjściu układu.

Jeśli na przykład gęstość widmowa mocy szumów wejściowych może być uważana w przedziale częstotliwości objętych charakterystyką filtru za wartość stałą μ , to średnia moc szumów wyjściowych wyniesie

$$\frac{\mu^2}{2\pi} \int_{-\infty}^{+\infty} |K(i\omega)|^2 d\omega.$$

Można wykazać, że dla wszystkich zakłóceń z ograniczoną powierzchnią μ , energia zakłóceń wyjściowych nie przekracza wartości określonej powyższym wzorem. Celowość operowania powierzchnią zakłóceń związana jest z tym, że przy działaniu na układ liniowy krótkich impulsów $z(t)$, reakcja jego prawie zupełnie nie zależy od kształtu sygnału, lecz jedynie od jego powierzchni.

Wyznamy teraz spośród wszystkich sygnałów zakłócających o ograniczonej powierzchni

$$\int_{-\infty}^{+\infty} |z(t)| dt = \mu, \quad (7)$$

takie, którym odpowiada maksymalna energia zakłóceń na wyjściu układu

$$E = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |Z(i\omega)K(i\omega)|^2 d\omega. \quad (8)$$

Tutaj

$$Z(i\omega) = \int_{-\infty}^{+\infty} z(t)e^{-i\omega t} dt, \quad K(i\omega) = \int_{-\infty}^{+\infty} k(t)e^{-i\omega t} dt.$$

Wykażemy, że są to sygnały o stałej gęstości widmowej. Dowód opiera się na nierówności

$$|Z(i\omega)| \leq \int_{-\infty}^{+\infty} |z(t)| \cdot |e^{-i\omega t}| dt = \int_{-\infty}^{+\infty} |z(t)| dt = \mu,$$

skąd dla energii zakłóceń otrzymujemy oszacowanie

$$E = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |ZK|^2 d\omega \leq \frac{\mu^2}{2\pi} \int_{-\infty}^{+\infty} |K(i\omega)|^2 d\omega = \text{const.}$$

Wobec tego największa wartość energii E jest osiągana dla sygnałów, przy których obie nierówności przechodzą w równość, skąd otrzymujemy

$$|Z(i\omega)| = \mu.$$

Sygnałem takiego typu jest na przykład kombinacja liniowa impulsów

$$z(t) = \sum_k c_k \delta(t - t_k), \quad \sum_k c_k = \mu,$$

poprzesuwanych w fazie.

Dla określonych powyżej najniekorzystniejszych zakłóceń energia na wyjściu wyrazi się wzorem

$$E = \frac{\mu^2}{2\pi} \int_{-\infty}^{+\infty} |K(i\omega)|^2 d\omega. \quad (9)$$

Obecnie sformułujemy ściśle rozpatrywany problem.

Niech będzie zadany sygnał wejściowy $e(t)$. Należy znaleźć układ odtwarzający z minimalnym odkształceniem (6) zadany przebieg $x(t)$, przy czym energia zakłóceń na wyjściu nie może przekraczać zadanej wartości E . Jeżeli zauważymy, że z pomocą związku $h'(t) = k(t)$ wiążącego wprowadzone funkcje przejścia, na mocy równości Plancherera możemy zapisać energię zakłóceń wzorem

$$E = \int_0^{\infty} h'^2(t) dt, \quad (10)$$

to zadanie nasze, z punktu widzenia matematycznego, jest zadaniem z rachunku wariacyjnego na minimum funkcjonału (6) przy izoperymetrycznym warunku ubocznym (10).

Przy podobnym podejściu do zagadnienia można wyznaczyć optymalną charakterystykę układu poddanego działaniu sygnału niestochastycznego o zadanym kształcie. Jeśli zaś sygnał działający na układ składa się ze składowej stochastycznej stacjonarnej e_s , niestochastycznej e_n i zakłóceń stochastycznych stacjonarnych z_s , to można wykazać [6], że średnia reakcja układu nie zależy od składowych stochastycznych, lecz określa się składową niestochastyczną e_n . Można więc znaleźć minimum błędu kwadratowego (6) I dla przebiegów niestochastycznych przy założonej wielkości błędu średniokwadratowego (3) $\bar{I} = \text{const}$ (w którym

uwzględnia się składowe stacjonarne) oraz przy założonej średniej mocy zakłóceń $E = \text{const}$ (lub też minimum $\bar{I}$ przy $I = \text{const}$, $E = \text{const}$).

Nieco inną metodę podano w pracy [6], gdzie składową niestochastyczną rozwija się w szereg w przedziale $[0, T]$

$$e_n(t - \tau) = e_n(t) - e'_n(t)\tau + \dots$$

i otrzymuje się odpowiednio

$$u_n(t) = \sum_{k=1}^n \mu_k e_k^{(k)}(t),$$

gdzie

$$\mu_k = \int_0^T \tau^k k(\tau) d\tau, \quad k = 0, 1, 2, \dots, n.$$

Przyrównując następnie $u_n(t)$ do zadanej funkcji czasu otrzymuje się warunki na μ_k . Wyznaczenie optymalnej funkcji przejściowej $k(t)$ sprowadza się tu do znalezienia minimalnej wartości funkcjonału $\bar{I}$ przy warunkach dodatkowych nałożonych na μ_k . Nie trudno zauważyć, że metoda powyższa komplikuje się w przypadku bardziej złożonych sygnałów (na przykład sygnałów, które nie mogą być przedstawione wielomianem algebraicznym dostatecznie niskiego rzędu).

Ponieważ zagadnienia związane z określeniem optymalnych charakterystyk układu przy sygnałach stochastycznych i stacjonarnych są wyczerpująco omówione w literaturze, zajmiemy się poniżej głównie wyznaczeniem charakterystyk układów poddanych działaniu sygnałów niestacjonarnych o zadanym kształcie.

2.1. RÓWNANIE CHARAKTERYSTYK OPTYMALNYCH

Zgodnie z metodą mnożników Lagrange'a równanie na funkcję $h(t)$ znajdziemy przyrównując do zera wariację funkcjonału pomocniczego

$$W = I + \lambda E, \quad (11)$$

w którym λ oznacza mnożnik Lagrange'a.

Jeśli $\eta(t)$ będzie dowolną funkcją znikającą poza skończonym przedziałem i równą zero dla $t=0$, wtedy warunek ekstremum funkcjonału (11) możemy napisać jako

$$\delta W = \left\{ \frac{d}{d\varepsilon} W[h + \varepsilon\eta] \right\}_{\varepsilon=0} = 0. \quad (12)$$

W dalszym ciągu obliczymy δW . Mamy

$$\frac{1}{2} \frac{dW}{d\varepsilon} = \frac{1}{2} \frac{dI}{d\varepsilon} + \frac{\lambda}{2} \frac{dE}{d\varepsilon}.$$

Za pomocą (3) i (4) od wyznaczonego sta możemy

$$\begin{aligned}\frac{1}{2} \frac{dI}{dt} &= - \int_0^{\infty} \sin \pi t \int_0^{\infty} d\epsilon e^{-\epsilon t} [1] u(t) - w(t) \\ \frac{1}{2} \frac{dK}{dt} &= - \int_0^{\infty} d\epsilon \sin \pi t \epsilon t\end{aligned}$$

a następnie

$$\frac{1}{2} \frac{dW}{dt} = - \int_0^{\infty} \sin \pi t \left[\epsilon t - \int_0^{\infty} d\epsilon e^{-\epsilon t} t [1] u(t) - w(t) \right].$$

Stosując z warunków granicznych (12) wykorzystując dowolność funkcji $\eta(t)$ otrzymujemy równanie

$$h(t) + \int_0^{\infty} d\epsilon e^{-\epsilon t} t [1] u(t) - w(t) = 0. \quad (13)$$

Stwierdzenie, że nasz sposób opisu ma charakterystykę $h(t)$. Przepiszemy je nieco inaczej, przy użyciu sta ϵ :

$$h(t) = \int_0^{\infty} d\epsilon t e^{-\epsilon t} \epsilon = t \int_0^{\infty} d\epsilon e^{-\epsilon t} \epsilon \sigma = t \quad (14)$$

Stąd

$$\int_0^{\infty} d\epsilon e^{-\epsilon t} \epsilon u(t) = \int_0^{\infty} d\epsilon h'(\epsilon) b(t - \epsilon)$$

I równanie (13) możemy przepisać w postaci

$$- h(t) + \int_0^{\infty} b(t - \tau) h'(\tau) d\tau = 0. \quad (15)$$

Obecnie dowiódzmy, że dla $\lambda \geq 1$ rozwiązanie równania (15) realizuje minimum funkcjonera (11). W istocie, ułożając drugą wariancję $\delta^2 W$ ze wzoru

$$\delta^2 W = \left\{ \frac{d^2}{d\epsilon^2} W[h + \epsilon \eta] \right\}_{\epsilon=0},$$

znajdujemy

$$\delta^2 W = \int_0^{\infty} \left(\frac{db}{d\epsilon} \right)^2 dt + \lambda \int_0^{\infty} \eta^2 dt.$$

gdzie

$$\tilde{u} = \frac{d}{dt} \int_0^{\infty} e(t-\tau) [h(\tau) + \varepsilon \eta(\tau)] d\tau.$$

Ponieważ $\lambda < 0$, więc $\delta^2 W > 0$, jak wiadomo, jest wystarczającym warunkiem minimum.

2.2. JEDNOZNACZNOŚĆ ROZWIĄZANIA RÓWNAŃ (15)

Gdy $k(t) = h'(t)$, wtedy zamiast (15) możemy napisać

$$-\lambda k'(t) + \int_0^{\infty} b(t-\tau) k(\tau) d\tau = c(t). \quad (16)$$

Równania (15) i (16) można sprowadzić do równań czysto całkowych, wykorzystując równości

$$k'(t) = \int_0^{\infty} \delta'(t-\tau) k(\tau) d\tau - k(0) \delta(t),$$

$$h''(t) = \int_0^{\infty} \delta''(t-\tau) h(\tau) d\tau - k(0) \delta(t).$$

Oznaczając

$$I_1(t) = -\lambda \delta'(t) + b(t), \quad I_2(t) = -\lambda \delta''(t) + b'(t),$$

zamiast (15) i (16) otrzymamy równania

$$\int_0^{\infty} I_1(t-\tau) k(\tau) d\tau = c(t) - \lambda k(0) \delta(t), \quad (15a)$$

$$\int_0^{\infty} I_2(t-\tau) h(\tau) d\tau + [b(+0) - b(-0)] h(t) = c(t) - \lambda k(0) \delta(t). \quad (16a)$$

Równania te są odpowiednio typu Focka i Wienera-Hopfa [3], [4]. W pracach tych autorów sformułowane są założenia, przy których równania takie przy zadanej prawej stronie mają jedno i tylko jedno rozwiązanie.

Prawe strony równań (15) i (16) zawierają niewiadomą wielkość $k(0)$, a więc twierdzeń tych nie możemy stosować bezpośrednio. Trzeba jednak zauważyć, że z uwagi na warunek izoperymetryczny (10) interesują nas jedynie rozwiązania, dla których $\int_0^{\infty} k^2 dt < +\infty$. Otóż można wykazać, że

zbieżność tej całki (i całki $\int_0^{\infty} k dt$) jednoznacznie wyznacza $k(0)$. Wobec tego możemy stwierdzić, że nasze równania mają co najwyżej jedno rozwiązanie w interesującej nas klasie funkcji (można wykazać, że wynika

to również z równania (12), bez powoływania się na teorię Wienera-Hopfa).

2.3. UKŁADY PRZEKAZUJĄCE

Zajmiemy się teraz przypadkiem, gdy $x(t)=e(t-t_0)$, to jest gdy sygnał, który z minimalnym zniekształceniem ma być odtworzony na wyjściu, jest sygnałem pierwotnym, opóźnionym o t_0 . Jak to wynika ze wzoru (14), mamy wtedy $c(t)=b(t-t_0)$, a równanie (16) przyjmie postać

$$-\lambda k'(t) + \int_0^{\infty} b(t-\tau) k(\tau) d\tau = b(t-t_0). \quad (17)$$

Rozwiązanie tego równania, spełniające warunek $\int_0^{\infty} k^2 dt < +\infty$, jest funkcją parametru t_0 . Zamiast (17) możemy więc napisać

$$-\lambda \frac{\partial}{\partial t} k(t, t_0) + \int_0^{\infty} b(t-\tau) k(\tau, t_0) d\tau = b(t-t_0). \quad (18)$$

Niech $k_0(t)=k(t, 0)$. Zatem $k_0(t)$ jest optymalną charakterystyką, odpowiadającą czasowi opóźnienia 0. Wykażemy, że charakterystyki $k(t, t_0)$ i $k_0(t)$ związane są równaniem

$$\frac{\partial}{\partial t} k(t, t_0) + \frac{\partial}{\partial t_0} k(t, t_0) + k(t, 0) \cdot k(t_0, 0) = 0 \quad (19)$$

Dla dowodu napiszmy równanie (18) raz dla $t_0=t_1$, drugi raz dla $t_0=t_2$; pierwsze z tych równań pomnożmy przez $k(t, t_2)$, drugie zaś przez $k(t, t_1)$, a następnie dodajmy i scałkujmy po t od zera do nieskończoności. Otrzymamy

$$\begin{aligned} \lambda k(0, t_1) k(0, t_2) + \frac{1}{2} \int_0^{\infty} \int_0^{\infty} dt d\tau \left\{ b(t-\tau) + b(\tau-t) \right\} \left\{ k(\tau, t_1) k(t, t_2) + k(t, t_2) k(\tau, t_1) \right\} = \\ = \int_0^{\infty} b(t-t_1) k(t, t_2) dt + \int_0^{\infty} b(t-t_2) k(t, t_1) dt. \end{aligned}$$

Następnie wykorzystamy wynikającą z (14) równość¹

$$b(t) + b(-t) = e^2(+\infty) = e^2_{-\infty}$$

oraz fakt, że, jak można udowodnić, mamy

$$\int_0^{\infty} k(t, t_0) dt = 1 \quad \text{dla } e_{\infty} \neq 0.$$

¹ Przy oczywistym założeniu $e(t)=0$, dla $t < 0$.

Wtedy otrzymamy

$$\begin{aligned} & \lambda k(0, t_1) k(0, t_2) + e_{\infty}^2 \int_0^{\infty} k(t, t_1) dt \int_0^{\infty} k(t, t_2) dt = \\ & = e_{\infty}^2 \int_0^{\infty} k(t, t_2) dt - \int_0^{\infty} b(t_1 - t) k(t, t_2) dt + e_{\infty}^2 \int_0^{\infty} k(t, t_1) dt - \int_0^{\infty} b(t_2 - t) k(t, t_1) dt. \end{aligned}$$

Całki z prawej strony można wyrugować za pomocą (18), przy czym skróćą się wyrazy zawierające e_{∞}^2 . Dzieląc przez λ otrzymujemy

$$\frac{\partial}{\partial t_1} k(t_1, t_2) + \frac{\partial}{\partial t_2} k(t_2, t_1) + k(0, t_1) k(0, t_2) = 0. \quad (20)$$

Można udowodnić (przy dodatkowych założeniach o zachowaniu się $k(t)$ w nieskończoności), że $k(t, t_0)$ jest symetryczną funkcją swoich argumentów

$$k(t, t_0) = k(t_0, t). \quad (21)$$

Ze związków (20) i (21) wynika (19).

Równanie (19) pozwala wyznaczyć $k(t, t_0)$, o ile znamy $k_0(t)$. Aby to wykazać, poddajemy (19) transformacji Laplace'a względem t i t_0 , oznaczając zmienną operatorową odpowiadającą t przez p , a zmienną operatorową odpowiadającą t_0 — przez q . Oznaczając dalej

$$\mathcal{L}_{t, t_0} \{k(t, t_0)\} = K(p, q); \mathcal{L}_t \{k_0(t)\} = K_0(p); \mathcal{L}_{t_0} \{k_0(t_0)\} = K_0(q)$$

w ten sposób z (19) otrzymamy równość

$$K(p, q) = \frac{K_0(p) + K_0(q) - K_0(p) \cdot K_0(q)}{p + q}.$$

Przez wzięcie transformatory odwrotnej względem q , dostajemy stąd, jeśli oznaczmy $K(p, t_0) = \mathcal{L} \{k(t, t_0)\}$, równanie

$$K(p, t_0) = K_0(p) e^{-pt_0} + [1 - K_0(p)] e^{-pt_0} \int_0^{t_0} k_0(\tau) e^{p\tau} d\tau. \quad (22)$$

Równanie (22) pozwala znaleźć $k(t, t_0)$ jeśli znamy $k(t, 0) = k_0(t)$.

2.4. CHARAKTERYSTYKI BEZ OPÓŹNIENIA

Zajmiemy się równaniem (18) dla zerowego czasu opóźnienia $t_0 = 0$. Wtedy

$$-\lambda k'_0(t) + \int_0^{\infty} b(t - \tau) k_0(\tau) d\tau = b(t). \quad (23)$$

Jak już pokazano, równanie to można sprowadzić do równania typu Wienera-Hopfa, dla którego podane są ogólne wzory na rozwiązanie. Tutaj

pokażemy, że gdy transformata Laplace'a funkcji $b(t)$, $B(p)$ jest funkcją wymierną, to istnieje rozwiązanie o postaci

$$k_0(t) = \sum_i A_i e^{a_i t} \quad \text{czyli} \quad K_0(p) = \sum_i \frac{A_i}{p - a_i}. \quad (24)$$

Powinniśmy mieć $\int_0^\infty k_0^2 dt < \infty$. Wynika stąd, że liczby a_i muszą mieć ujemne części rzeczywiste. Wyznamy liczby a_i i A_i . Podstawiając (24) do równania (23) i transformując otrzymaną równość według Laplace'a, po uporządkowaniu otrzymamy równanie

$$\sum_i \frac{A_i}{p - a_i} - [-\lambda a_i + B(p) + V_i] = B(p), \quad (25)$$

gdzie

$$V_i = \int_0^\infty b(-\sigma) e^{a_i \sigma} d\sigma = -\frac{e_\infty^2}{a_i} - B(-a_i).$$

Mnożąc równanie (25) przez $p - a_k$ i zakładając $p = a_k$ otrzymujemy

$$-\lambda a_k + B(a_k) + V_k = 0$$

lub

$$-\lambda a_k + B(a_k) - \frac{e_\infty^2}{a_k} - B(-a_k) = 0.$$

A więc liczby a_i są pierwiastkami o ujemnej części rzeczywistej równania charakterystycznego

$$\lambda p + \frac{e_\infty^2}{p} + B(-p) - B(p) = 0. \quad (26)$$

Aby teraz wyznaczyć liczby A_i , porównujemy residua lewej i prawej strony (25). Jeśli przez β_k oznaczymy bieguny funkcji $B(p)$ (a można wykazać, że jest ich akurat tyle, ile ujemnych pierwiastków równania 26), to dostaniemy

$$\sum_i \frac{A_i}{(\beta_k - a_i)} = 1, \quad (27a)$$

jeżeli biegun β_k jest pojedynczy oraz

$$\sum_i \frac{A_i}{(\beta_k - a_i)^{n_k}} = 0, \quad (27b)$$

dla bieguna o krotności $n_k > 1$.

Powyższe równania ze względu na A_i można rozwiązać w postaci zamkniętej posługując się pomocniczymi funkcjami

$$F(p) = \prod_i (p - a_i); \quad G(p) = \prod_k (p - \beta_k)^{n_k}. \quad (28)$$

Mamy mianowicie

$$A_i = - \frac{G(a_i)}{F'(a_i)}, \quad (29)$$

dla $K_0(p)$ zaś

$$K_0(p) = 1 - \frac{G(p)}{F(p)}. \quad (30)$$

Mając $K_0(p)$ znajdujemy $K(p, t_0)$ ze wzoru (22).

Na koniec można wykazać, że przy a_i i A_i wyznaczonych z (26) i (29) równanie (25) jest tożsamościowo spełnione, a tym samym wzory (24) dają rzeczywiście rozwiązanie równania (23).

2.5. PRZYKŁADY

Dla przykładu znajdziemy optymalne charakterystyki dla sygnału

$$e(t) = e^{-t}, \quad x(t) = e(t - t_0).$$

Ze wzoru (14) otrzymujemy

$$b(t) = \frac{1}{2} e^{-t}, \quad B(p) = \frac{1}{2} \frac{1}{1+p}, \quad \beta = -1.$$

Równanie charakterystyczne (26) ma postać

$$\lambda p + \frac{p}{1-p^2} = 0$$

i jeden pierwiastek ujemny $\alpha = -\sqrt{1 + \frac{1}{\lambda}}$. Z (28) obliczamy pomocnicze funkcje F i G jako

$$F = p + \sqrt{1 + \frac{1}{\lambda}}; \quad G = p + 1.$$

W końcu z (29) otrzymujemy

$$A = \sqrt{1 + \frac{1}{\lambda}} - 1$$

i z (24)

$$k_0(t) = A e^{\alpha t} = \left(\sqrt{1 + \frac{1}{\lambda}} - 1 \right) e^{-\sqrt{1 + \frac{1}{\lambda}} t},$$

$$K_0(p) = \frac{A}{p - \alpha} = \frac{\sqrt{1 + \frac{1}{\lambda}} - 1}{p + \sqrt{1 + \frac{1}{\lambda}}}.$$

Wzór ten podaje optymalne charakterystyki zerowe. Następnie z (22) znajdujemy

$$K(p, t_0) = \frac{A}{p^2 - a^2} \left\{ e^{at_0} (p+1) + e^{-pt_0} (a-1) \right\}.$$

skąd

$$k(t, t_0) = \begin{cases} A e^{at_0} \left(\operatorname{ch} at + \frac{1}{a} \operatorname{sh} at \right) & \text{dla } t \leq t_0, \\ A e^{at} \left(\operatorname{ch} at_0 + \frac{1}{a} \operatorname{sh} at_0 \right) & \text{dla } t > t_0. \end{cases}$$

Są to charakterystyki optymalne dla czasu opóźnienia t_0 .

Jako drugi przykład wyznaczmy charakterystyki optymalne dla sygnału $e(t) = 1(t)$, $x(t) = 1(t - t_0)$.

Ze wzoru (14) mamy

$$b(t) = 1(t),$$

zatem

$$\mathcal{L}\{b(t)\} = B(p) = \frac{1}{p}.$$

Równanie charakterystyczne (26) ma kształt następujący

$$\lambda p - \frac{1}{p} = 0,$$

oraz pierwiastek

$$a = -\sqrt{\frac{1}{\lambda}}.$$

Zatem funkcje pomocnicze (28) przyjmują postać

$$F = p - a, \quad G = p,$$

zaś

$$A = -\frac{a}{\lambda} = \sqrt{\frac{1}{\lambda}}.$$

Korzystając ze wzorów (30) i (22) otrzymujemy

$$K_0(p) = \frac{A}{p+A},$$

$$K(p, t_0) = \frac{A}{p^2 - A^2} \left\{ p e^{-At_0} - A e^{-pt_0} \right\}.$$

Wobec tego

$$k(t, t_0) = \begin{cases} A e^{-At_0} \operatorname{ch} At, & t \leq t_0 \\ A e^{-At} \operatorname{ch} At_0, & t > t_0. \end{cases} \quad (31)$$

Charakterystyka o identycznej postaci została wyznaczona w pracy [2].

Charakterystyka odpowiadająca sygnałowi jednostkowemu jest bardzo interesująca z praktycznego punktu widzenia, gdyż — jak to zaraz wykażemy — skok jednostkowy jest sygnałem najbardziej niekorzystnym z punktu widzenia zniekształceń sygnału (w metryce przestrzeni C) z pomiędzy wszystkich sygnałów z ograniczoną amplitudą.

W tym celu założymy, że sygnał $e(t)$ działa na układ liniowy, którego reakcja na impuls jednostkowy jest $k(t)$. Sygnał na wyjściu układu wyniesie

$$u(t) = \int_0^{\infty} e(t-\tau) k(\tau) d\tau. \quad (32)$$

W przypadku, gdy układ nie wprowadza zniekształceń

$$k(t) = k_0 \delta(t - t_0), \quad (33)$$

to znaczy, że odpowiedź układu na sygnał $\delta(t)$ różni się od tego sygnału tylko opóźnieniem t_0 i tłumieniem k_0 .

Podstawiając (33) do (32) otrzymujemy

$$u(t) = \int_0^{\infty} e(t-\tau) k_0 \delta(\tau - t_0) d\tau = k_0 e(t - t_0). \quad (34)$$

W przypadku układu, który wprowadza zniekształcenia mamy $u(t) \neq k_0 e(t - t_0)$. Zniekształcenia te można wyrazić za pomocą wzoru

$$\Delta = \max_t |u(t) - k_0 e(t - t_0)|. \quad (35)$$

Podstawiając wzory (32), (34) do (35) możemy napisać

$$\begin{aligned} \Delta &= \max \left| \int_0^{\infty} e(t-\tau) k(\tau) d\tau - \int_0^{\infty} k_0 e(t-\tau) \delta(\tau - t_0) d\tau \right| = \\ &= \max \left| \int_0^{\infty} e(t-\tau) [k(\tau) - k_0 \delta(\tau - t_0)] d\tau \right|. \end{aligned} \quad (36)$$

Sygnały działające na układ posiadają zwykle ograniczoną amplitudę i dlatego możemy założyć

$$|e(t)| \leq E_m.$$

Wyznamy teraz spośród sygnałów z ograniczoną amplitudą taki sygnał, który powoduje maksymalne zniekształcenie Δ . Wyrażenie (36) uzyskuje maksymalną wartość wtedy, gdy funkcja podcałkowa będzie stałe dodatnia. Wynika stąd, że zniekształcenie (36) maksymalną wartość uzyskuje w przypadku sygnału, który ma stałą amplitudę $\pm E_m$ i zmienia znak w zerach funkcji $k(\tau) - k_0 \delta(\tau - t_0)$. Dla układu, w którym funkcja ta nie posiada zer dla $t > 0$, najbardziej niekorzystnym sygnałem, ze wszystkich sygnałów z ograniczoną amplitudą, będzie sygnał o kształcie skoku

$E_m 1(t)$. Rezultat powyższy uzasadnia celowość stosowania sygnału jednostkowego jako sygnału wzorcowego przy badaniu aparatury elektrycznej.

3. SYGNAŁY MODULOWANE

Sygnały elektryczne stosowane w praktyce występują najczęściej w postaci modulowanej

$$e(t) = e_0(t) \cos \omega_0 t = \mathcal{R}_e \{e_0(t) e^{i\omega_0 t}\},$$

gdzie $e_0(t)$ określa charakter zmian obwiedni sygnału modulowanego, natomiast $\cos \omega_0 t = \mathcal{R}_e \{e^{i\omega_0 t}\}$ — reprezentuje tak zwaną falę nośną.

Ponieważ zwykle interesuje nas obwiednia sygnału modulowanego (fala nośna zostaje wyeliminowana w detektorze) wyznaczmy obwiednią sygnału na wyjściu układu liniowego.

Wyznamy najpierw reakcję układu na sygnał

$$e_0(t) e^{i\omega_0 t}.$$

Widmo tego sygnału wyraża się wzorem

$$E(i\omega) = E_0[i(\omega - \omega_0)],$$

gdzie

$$E_0(i\omega) = \mathcal{F}\{e_0(t)\},$$

sygnał zaś na wyjściu układu wyraża się jako

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} E_0[i(\omega - \omega_0)] K(i\omega) e^{i\omega t} d\omega.$$

Zatem reakcja układu na sygnał

$$e(t) = \mathcal{R}_e \{e_0(t) e^{i\omega_0 t}\}$$

ze względu na liniowość układu wyniesie

$$u(t) = \mathcal{R}_e \left\{ \frac{1}{2\pi} \int_{-\infty}^{+\infty} E_0[i(\omega - \omega_0)] K(i\omega) e^{i\omega t} d\omega \right\}. \quad (37)$$

Jeśli sygnał modulowany działa na układ o własnościach selektywnych (na przykład wzmacniacz rezonansowy nastrojony na częstotliwość nośną) wygodniej jest posługiwać się pojęciem funkcji przenoszenia, w której częstotliwość ω odmierza się od częstotliwości rezonansowej ω_0 , to jest

$$K(i\omega) = K_{ob}[i(\omega - \omega_0)] = K_{ob}(i\Omega). \quad (38)$$

Podstawiając (38) do wzoru (37) otrzymujemy

$$u(t) = \mathcal{R}_e \left\{ \frac{e^{i\omega_0 t}}{2\pi} \int_{-\infty}^{+\infty} E_0(i\Omega) K_{ob}(i\Omega) e^{i\Omega t} d\Omega \right\} = u_0(t) \cos \omega_0 t,$$

gdzie

$$u_0(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} E_0(i\Omega) K_{ob}(i\Omega) e^{i\Omega t} d\Omega. \quad (39)$$

Wzór (39) możemy również wyrazić za pomocą spłotu obwiedni sygnału wejściowego $e(t)$ z funkcją przejściową obwiedni

$$h_{ob}(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{K_{ob}(i\Omega)}{i\Omega} e^{i\Omega t} d\Omega,$$

a mianowicie

$$u_0(t) = \frac{d}{dt} \int_0^{\infty} e_0(t-\tau) h_{ob}(\tau) d\tau.$$

Wzór ten ma analogiczną postać do stosowanych w rozdziałach 2 i 3 wzorów na spłot sygnału $e(t)$ z funkcją przejściową $k(t)$. Wynika stąd, że optymalne charakterystyki częstotliwościowe wyznaczone metodą podaną w rozdziałach 2.3, 2.4, 2.5 i odnoszące się do układów bez modulacji (tzw. videoczęstotliwości) mogą być w prosty sposób przetransformowane (przez translację częstotliwości $\Omega = \omega - \omega_0$) na optymalne charakterystyki układów z modulacją. Przy transformacji tej kształt charakterystyki nie ulega zmianie.

4. FILTRY „IDEALNE“

W urządzeniach telekomunikacyjnych z tak zwanym częstotliwościowym rozdzieleniem sygnałów stosuje się filtry, które przepuszczają tylko określone (przez normy) pasmo częstotliwości i obcinają częstotliwości leżące poza pasmem przepuszczanym². Przy projektowaniu takich filtrów stykamy się z koniecznością wyznaczenia charakterystyk zapewniających minimalne zniekształcenia sygnałów przy ograniczonym wpływie zakłóceń.

Obecnie wyznaczymy optymalne charakterystyki filtra *idealnego* zapewniające przeniesienie sygnału wzorcowego $1(t)$ z minimalnymi zniekształceniami

$$I = \int_{-\infty}^{+\infty} [h_{ob}(t) - 1(t-t_0)]^2 dt \quad (40)$$

² Ściśle biorąc, układy te nie mogą być zrealizowane, gdyż charakterystyki ich nie spełniają na ogół warunku $k(t)=0$ dla $t<0$. Jednakże charakterystyki układów selektywnych mogą być z wystarczającą dla praktyki dokładnością aproksymowane za pomocą charakterystyk dających się realizować

przy warunku dodatkowym (9) $E = \text{const}$ oraz przy założeniu, że widmo sygnału wyjściowego znika przy $|\Omega| > \Omega_0$.

We wzorze (40) za dolny przedział całkowania przyjęliśmy $-\infty$, gdyż sygnał $h_{ob}(t)$ dla filtru idealnego może nie zniknąć przy $t < 0$. Ponieważ widmo sygnału wyjściowego

$$\frac{K_{ob}(i\Omega)}{i\Omega} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} h_{ob}(t) e^{-i\Omega t} dt,$$

w myśl założenia równa się zeru na zewnątrz pasma częstotliwości przepuszczanych przez filtr idealny, możemy napisać

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} h_{ob}(t) e^{-i\Omega t} dt = 0, \quad \text{dla } |\Omega| > \Omega_0, \quad (41)$$

przy czym Ω_0 jest górną częstotliwością pasma przepuszczanego.

Zagadnienie polega zatem na wyznaczeniu funkcji $h(t)$, która zapewnia minimum funkcjonału

$$H_1 = I + \lambda_1 E,$$

przy czym na możliwe rozwiązanie nakłada się dodatkowy warunek (41).

Zagadnienie powyższe jest zagadnieniem wariacyjnym na ekstremum warunkowe. Zagadnienie takie można rozwiązywać metodą Lagrange'a, zwaną też metodą nieokreślonego mnożnika funkcyjnego.

Zgodnie z powyższą metodą, pomnożmy wyrażenie (41) przez nieokreśloną na razie funkcję $\lambda(\Omega)$ i scałkujmy otrzymaną funkcję w przedziale, który leży na zewnątrz pasma częstotliwości przepuszczanych.

Otrzymamy wtedy zamiast (41) warunek następujący

$$D = \int_{-\infty}^{-\Omega_0} \left[\frac{\lambda(\Omega)}{2\pi} \int_{-\infty}^{+\infty} h_{ob}(t) e^{-i\Omega t} dt \right] d\Omega + \int_{\Omega_0}^{+\infty} \left[\frac{\lambda(\Omega)}{2\pi} \int_{-\infty}^{+\infty} h_{ob}(t) e^{-i\Omega t} dt \right] d\Omega = 0, \quad |\Omega| > \Omega_0.$$

Zmieniając we wzorze powyższym kolejność całkowania otrzymujemy

$$D = \frac{1}{2\pi} \int_{-\infty}^{+\infty} h_{ob}(t) \left[\int_{-\infty}^{-\Omega_0} \lambda(\Omega) e^{-i\Omega t} d\Omega + \int_{\Omega_0}^{+\infty} \lambda(\Omega) e^{-i\Omega t} d\Omega \right] dt = 0, \quad |\Omega| > \Omega_0.$$

W myśl metody Lagrange'a znajdujemy teraz ekstremum bezwzględne funkcjonału

$$H = H_1 + D = I + D + \lambda_1 E. \quad (42)$$

Przyrównując do zera pierwsze wariacje względem $h_{ob}(t)$ oraz $\lambda(\Omega)$ wyrażenia (42), otrzymujemy następujące równania:

$$-\lambda_1 h''_{ob} + h_{ob} - 1(t - t_0) + \frac{1}{4\pi} \left\{ \int_{-\infty}^{-\Omega_0} \lambda(\Omega) e^{-i\Omega t} d\Omega + \int_{\Omega_0}^{\infty} \lambda(\Omega) e^{-i\Omega t} d\Omega \right\} = 0, \quad (43a)$$

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} h_{ob}(t) e^{-i\Omega t} dt = 0, \quad \text{dla } |\Omega| > \Omega_0. \quad (43b)$$

Poddając równanie (43a) transformacji Fouriera oraz uwzględniając (43b) otrzymujemy

$$\frac{K_{ob}(i\Omega)}{i\Omega} \left[-\lambda_1 (i\Omega)^2 + \lambda_1 h'(0) + 1 \right] = \frac{e^{-i\Omega t_0}}{i\Omega} \quad \text{dla } |\Omega| < \Omega_0.$$

Stąd oraz z (43b) otrzymujemy wreszcie

$$K_{ob}(i\Omega) = \begin{cases} \frac{e^{-i\Omega t_0}}{-\lambda_1 (i\Omega)^2 + \lambda_1 h'(0) + 1}, & |\Omega| < \Omega_0, \\ 0, & |\Omega| > \Omega_0. \end{cases}$$

We wzorze tym $h'_{ob}(0)$ może być określone na przykład z równości

$$h'_{ob}(0) = \int_{-\Omega_0}^{+\Omega_0} K_{ob}(i\Omega) d\Omega,$$

λ_1 zaś z warunku izoperymetrycznego (9).

Dla $\lambda_1 = 0$ otrzymujemy charakterystykę [2]

$$K_{ob}(i\Omega) = \begin{cases} e^{-i\Omega t} & \text{dla } |\Omega| < \Omega_0, \\ 0 & \text{dla } |\Omega| > \Omega_0. \end{cases}$$

Charakterystyce tej odpowiada prostokątna charakterystyka amplitudy

$$A(\Omega) = |K_{ob}(i\Omega)| = \begin{cases} 1 & \text{dla } |\Omega| < \Omega_0, \\ 0 & \text{dla } |\Omega| > \Omega_0, \end{cases}$$

oraz liniowa charakterystyka fazowa.

Charakterystyce tej odpowiada

$$h_{ob}(t) = \frac{1}{2} + \frac{1}{\pi} \text{Si } \Omega_0 (t - t_0),$$

gdzie Si oznacza sinus całkowity. Charakterystyka powyższego typu, mimo że należy do charakterystyk nie dających się zrealizować, może być z dostateczną dla praktyki dokładnością aproksymowana przez charakterystyki układów możliwych do realizacji.

Zakładając na przykład, że funkcja tłumienia układu wyraża się wzorem

$$\frac{1}{K_{ob}(ix)} = \prod_{k=1}^n (ix - x_k), \quad x = \frac{\Omega}{\Omega_0}^3, \quad (44)$$

możemy wyznaczyć takie rozstawienie zer x_k -tej funkcji, które zapewnia najlepsze przybliżenie (w sensie Czebyszewa) charakterystyki idealnej

$\frac{1}{|K_{ob}(ix)|} = 1$ w przedziale $(-1, +1)$ oraz dostatecznie duże tłumienie poza pasmem przepuszczania ($x > 1$). W tym celu założmy, że kwadrat modułu funkcji (44) jest proporcjonalny do

$$A_0 + L T_{2n}(x),$$

gdzie $T_{2n}(x)$ — oznacza wielomian Czebyszewa $2n$ -tego rzędu,

L — maksymalne odchylenie tego wielomianu od $A_0 = \text{const}$

Jeśli nierównomierność tłumienia w pasmie przepuszczania układu wynosi a dB (rys. 1a), to mamy

$$A_0 = \frac{e^{2a} + 1}{2}, \quad L = \frac{e^{2a} - 1}{2} \quad (\text{rys. 1b}).$$

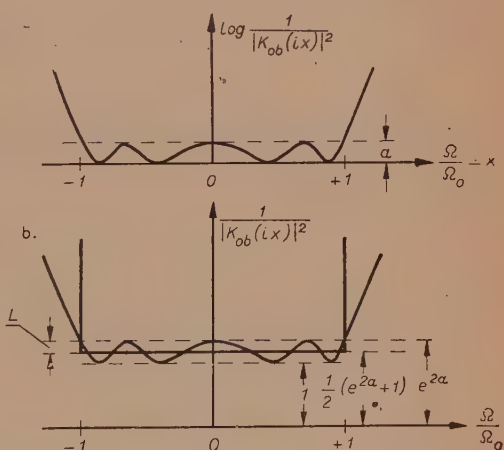
Wprowadzając współczynnik tłumienia

$$T = \sqrt[2n]{\frac{e^a - 1}{e^a + 1}},$$

który równa się zeru dla $a=0$ i dąży do jedności przy bardzo dużym a ; wielomian (44) można wyrazić jako

$$\frac{1}{(1 - T^{2n})^2} [1 + T^{4n} + 2T^{2n}T_{2n}(x)]. \quad (45)$$

Aby wielomian ten był proporcjonalny do wielomianu (44), zera ich powinny się pokrywać. Zera te będą leżały w ogólnym przypadku na płaszczyźnie zmiennej, zespolonej $x = z + iy$. Wprowadzając współrzędne eliptyczne według wzorów



Rys. 1. Aproksymacja charakterystyki filtru idealnego.

³ Funkcje przenoszenia takiego typu posiadają na przykład układy złożone ze wzmacniaczy rezonansowych, zawierających jeden lub dwa obwody w każdym stopniu.

$$z = \frac{1}{2} \left(\varrho + \frac{1}{\varrho} \right) \cos \varphi,$$

$$y = \frac{1}{2} \left(\varrho - \frac{1}{\varrho} \right) \sin \varphi,$$

$$x = \frac{1}{2} \left(\varrho e^{i\varphi} + \frac{1}{\varrho} e^{-i\varphi} \right),$$

do wzoru (45) otrzymujemy ⁴

$$\frac{1}{(1+T^{2n})^2} \left[1 + T^{4n} + T^{2n} \left(\varrho^{2n} e^{i2n\varphi} + \frac{1}{\varrho^{2n}} e^{-i2n\varphi} \right) \right].$$

Nie trudno teraz zauważyć, że funkcja powyższa równa się zeru dla

$$\varrho = \frac{1}{T}, \quad 2n\varphi = (2k-1)\pi, \quad k=1, 2, \dots, 2n.$$

wynika stąd, że zera wielomianu (44) powinny leżeć na elipsie w punktach o współrzędnych

$$z_k = \frac{1}{2} \left(T + \frac{1}{T} \right) \cos \frac{2k-1}{2n} \pi,$$

$$y_k = \frac{1}{2} \left(T - \frac{1}{T} \right) \sin \frac{2k-1}{2n} \pi.$$

Przyjmując zatem określoną wartość tłumienia a , możemy ze wzorów powyższych wyznaczyć optymalne położenie biegunów x_k funkcji przenoszenia $K_{ob}(ix)$, a następnie określić optymalne parametry układu projektowanego.

WYKAZ LITERATURY

1. Dwork B.: *Detection of a pulse superimposed on fluctuation noise*. P. I. R. E. 1950, s. 771.
2. Kulikowski R., Plebański J.: *Optimum characteristics of linear puls systems*. Bull. de l'Acad. Pol. des Sc., nr 1, 1955.
3. Smirnow W. I.: *Kurs wysszej matematiki*. T. IV, s. 179.

⁴ Wzór ten wynika z następujących zależności:

$$T_{2n} = \frac{1}{2} [(x + \sqrt{x^2 - 1})^{2n} + (x - \sqrt{x^2 - 1})^{2n}],$$

$$\sqrt{x^2 - 1} = \frac{1}{2} \left(\varrho e^{i\varphi} - \frac{1}{\varrho} e^{-i\varphi} \right),$$

$$x + \sqrt{x^2 - 1} = \varrho e^{i\varphi}, \quad x - \sqrt{x^2 - 1} = \frac{1}{\varrho} e^{-i\varphi}.$$

4. Wiener N., Hopf: *Über eine Klasse singularer Integralgleichungen*. Preus. Akad., 1931.
5. Wiener N.: *Extrapolation, interpolation and smoothing of stationary time series*. M. I. T., 1950.
6. Zadeh L. A., Ragazzini J. R.: *An extension of Wiener's theory of prediction*. Jour. of Appl. Physics. v. 2, nr 7, 1950.

Р. КУЛИКОВСКИ, А. РЫБАРСКИ

КРИТЕРИЙ КВАДРАТИЧЕСКИХ ИСКАЖЕНИЙ ПРИ ОГРАНИЧЕННОЙ МОЩНОСТИ ПОМЕХ

Резюме

В качестве критерия искажений сигнала $e(t)$ в линейной системе с переходной функцией $h(t)$ принимается выражение I (6), где $x(t)$ — линейный функционал $e(t)$. Если спектральную плотность входных шумов можно считать в полосе пропускания фильтра величиной постоянной μ , то средняя мощность выходных шумов выражается формулой (9). Показано также, что среди всех помех с ограниченной поверхностью μ (7), энергия помех на выходе фильтра не превышает величины (9).

Вопрос сводится затем к определению минимального значения искажений (6) при добавочном условии $E = \text{const}$ (9), (10). Приравнявая первую вариацию функционала $W = I + \lambda E$ по $h(t)$ к нулю получаем уравнение (15), однозначность решений которого исследована в главе 3.2. В главе 3.3 приводится решение уравнения (15), определяющее оптимальную переходную функцию системы. Затем рассмотрено конкретные примеры при входных сигналах

$$e(t) = e^{-t}, \quad x(t) = e(t - t_0) \quad \text{и} \quad e(t) = 1(t), \quad x(t) = 1(t - t_0).$$

В главе 4 показано, что полученные результаты можно перенести при помощи трансляции частоты на случай амплитудно-модулированных сигналов.

Затем определено характеристики системы, обеспечивающей минимальные искажения единичного сигнала при $E = \text{const}$ и добавочном условии, которое сводится к требованию, чтобы частотные характеристики системы исчезали вне полосы пропускания. В случае $E = 0$ оптимальная характеристика амплитуды имеет форму прямоугольника, который, как это показано, можно аппроксимировать при помощи функции затухания (44) с любой точностью.

R. KULIKOWSKI A. RYBARSKI

CRITERION FOR ROOT SQUARE DISTORTION AT LIMITED NOISE POWER

Summary

As a criterion for distortion of the signal $e(t)$ in a linear system having a transfer function $h(t)$, we take expression $I(6)$, where $x(t)$ is a linear operation of the function $e(t)$. If the spectral density of the input noise power to have a constant value μ in the frequency interval included in the filter characteristic, then the average output noise power will be equal to (9). It is also shown that among all disturbances with a limited area μ . (7), the output noise energy does not exceed the value given by formula (9).

The problem involves determining the minimum value of the distortion (6) at the subsidiary condition that $E = \text{const}$ (9), (10). To first variation of $W = I + \lambda E$ with respect to $u(t)$ being set equal to zero, we obtain Eq. (15). The question of whether the solution of Eq. (15) is single-valued, is considered under point 3.2. The solution of Eq. (15) defining the optimum transfer function of the system is given under point 3.3. Next, specific examples with input signals $e(t) = e^{-t}$, $x(t) = e(t - t_0)$ and $e(t) = 1(t)$, $x(t) = 1(t - t_0)$ are discussed.

In section 4, it is shown that the results presented in the previous sections can be applied to amplitude modulated signals by a translation of frequency.

At last, the authors determine the characteristics of a system which distorts, to a minimum extent, a unit signal with $E = \text{const}$ and whose frequency characteristic vanishes outside a defined frequency band. In the case of $E = 0$, the optimum amplitude characteristic has a rectangular shape. It is shown that this shape can be approximated to any desired accuracy with the help of the damping function (44).

621.3.09 : 536.758

J. SEIDLER

O pojęciu miary nieokreśloności [entropii]

Rękopis dostarczono 16. 12. 1955

Entropia i miara ilości informacji mają bezpośrednie zastosowanie do charakterystyki systemów ziarnistych, gdy wymaga się, aby prawdopodobieństwo błędów było równe zeru. Celem pracy jest ustalenie takich własności entropii, które pozwalałyby ją stosować jako charakterystykę systemów faktycznie ciągłych oraz systemów dopuszczających małe, ale skończone prawdopodobieństwo błędów. Wprowadzono ogólne pojęcie liczbowej charakterystyki systemu tak zwaną miarę rozrzutu. Wykazano, że entropia ma sens takiej właśnie wielkości i ustalono jej powiązanie z innymi typowymi miarami rozrzutu.

1. WSTĘP

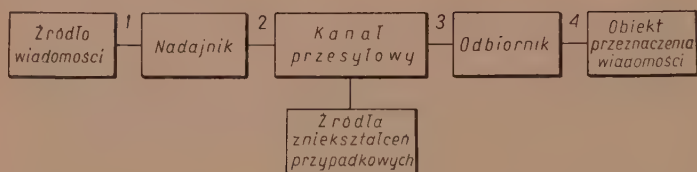
Pojęcia entropii zbioru sygnałów i miary ilości informacji, wprowadzone do teorii telekomunikacji w pracy [10], mają stosunkowo prosty sens dla zbiorów ziarnistych sygnałów.

Gdy parametry określające sygnał mogą zmieniać się w sposób ciągły, entropie, *a priori* i *a posteriori* obliczone w sposób formalnie podobny jak dla rozkładów ziarnistych, pozwalają uzyskać asymptotyczne wyrażenie dla ilości sygnałów, które mogą być bezbłędnie przesyłane. Aby zbliżyć się do takich warunków przesyłania, trzeba przeprowadzić kodowanie. Tym samym jednak system ciągły zostaje faktycznie zastąpiony systemem ziarnistym. Tymczasem bardzo często w praktyce są używane systemy nieskwantyzowane i wcale do skwantyzowania nieprzewidziane. Nasuwa się pytanie czy entropii można nadać inną interpretację, aniżeli wielkości opisującej potencjalne możliwości systemu w razie zakodowania, a tym samym używać jej jako charakterystyki systemów faktycznie ciągłych.

Wspomniana prosta interpretacja entropii zbiorów ziarnistych dotyczy nierealnego przypadku, gdy długość ciągów sygnałów wzrasta nieograniczenie, a prawdopodobieństwo błędów dąży do zera. W praktyce, szczególnie tego drugiego warunku ani nie można, ani na ogół nie trzeba realizować. Wyłania się więc inny problem, mianowicie czy entropia zbiorów ziarnistych może być użyta jako praktyczna charakterystyka syste-

mów posługujących się sygnałami o skończonej długości i dopuszczających skończone prawdopodobieństwo błędów.

Zanim rozpatrzmy te zagadnienia, zajmiemy się ogólną analizą telekomunikacji w obecności zakłóceń. Posługiwać się będziemy schematem przedstawionym na rys. 1.



Rys. 1. Schemat systemu komunikacyjnego.

1. wiadomość, 2. sygnał nadany — x , 3. sygnał odebrany — y , 4. sygnał odtworzony $x^*(y)$.

W czasie przesyłania sygnały ulegają z zasady zniekształceniom i odtworzona na podstawie odebranego sygnału wiadomość różni się na ogół od wiadomości nadanej.

Zniekształcenia są ostatecznym czynnikiem ograniczającym zasięg, na jaki można przekazywać wiadomości. Z punktu widzenia obiektu odbierającego wiadomości można wyróżnić dwa zasadnicze typy zniekształceń:

I. zniekształcenia spowodowane przez czynniki w pełni znane z punktu widzenia urządzenia odbiorczego,

II. zniekształcenia spowodowane przez czynniki przypadkowe; znane tylko regularności zjawisk masowych, które będziemy nazywali własnościami statystycznymi zakłóceń.

Jako przykłady zniekształceń typu I przytoczymy:

- 1) znane osłabienie sygnału w trakcie przesyłania,
- 2) nałożenie się na sygnał innego przebiegu o znanym kształcie (na przykład przydźwięku sieciowego).

Przykładami zniekształceń typu II są:

- 1) zniekształcenie spowodowane przypadkowymi wahaniem stanu jonosfery,
- 2) nałożenie się szumu, to jest przebiegu, którego dokładnie z punktu widzenia odbiorczego nie znamy, ale znamy własności statystyczne, a więc na przykład: wartość średnią, średni rozrzut kwadratowy.

Korekcja zniekształceń typu I, poza nierealnymi przypadkami, jest w zasadzie możliwa. W podanych przykładach może być ona uzyskana przez zastosowanie odpowiedniego wzmacniacza lub układu kompensacyjnego.

Obecność zniekształceń typu II jest nieuniknioną konsekwencją praw fizycznych. Zawsze obecne są efekty fluktuacji termicznych. W wielu

zagadnieniach praktycznych istotną rolę odegrają znacznie silniejsze zakłócenia atmosferyczne, przemysłowe lub też zakłócenia pochodzące od innych systemów komunikacyjnych. Znajomość statystycznych własności zakłóceń pozwala na zmniejszenie ich wpływu, ale w zasadzie nigdy na ich całkowite wyeliminowanie. Pojęcie komunikacji musi być poddane rewizji. Zamiast wiadomości bezbłędnie odtwarzanej przez odbiorcę musimy się zadowolić wiadomością najbardziej *pewną* w sensie tego czy innego kryterium. Kryteria te mają charakter obiektywny, przy czym bardzo często wiadomość wysoce *pewną* można traktować w praktyce jako faktycznie nadaną.

Podstawowym narzędziem do traktowania problemów komunikacji w takim ujęciu jest rachunek prawdopodobieństwa. Teoria komunikacji staje się więc w zasadzie zagadnieniem statystyki matematycznej; ma jednak ona swoją specyfikę. Z punktu widzenia urządzenia odbiorczego sygnały nadawane trzeba traktować jako przypadkowe. Dokładniejsza analiza struktury i własności źródła wiadomości pozwala na ustalenie statystycznych własności nadawanych wiadomości. Gdy analiza taka nie została przeprowadzona lub jest zbyt skomplikowana, własności statystyczne można ustalić doświadczalnie. Ostatecznie, w przeciwieństwie do wielu zagadnień statystyki, w przypadku komunikacji mogą być wyraźnie ustalone prawa rozkładu *a priori*.

Drugą cechą charakterystyczną, na którą zwrócimy uwagę, jest specyficzna budowa wiadomości i sygnałów. Bardzo często wiadomości lub sygnały mają strukturę ciągów zbudowanych z elementarnych składników. Jako typowy przykład takich wiadomości podamy drukowane teksty, które są ciągami liter.

Z kolei przystąpimy do sprecyzowania pojęcia komunikacji w obecności przypadkowych zniekształceń oraz określimy miarę stopnia *pewności* odtworzonej wiadomości.

2. CHARAKTERYSTYKI LICZBOWE SYSTEMÓW KOMUNIKACYJNYCH

Zajmiemy się wpierw miarą liczbową zniekształceń. Jeżeli w konkretnych przypadkach wziąć pod uwagę sposób, w jaki obiekt dla którego są przeznaczone wiadomości reaguje na otrzymywane wiadomości, jak je wykorzystuje, to na ogół udaje się wprowadzić liczbową miarę zniekształceń. Można geometrycznie zapatrywać się na nią jako na miarę odległości sygnałów i żądać, aby wielkość ta posiadała dodatkowo własności miary odległości.

Oto kilka przykładów miary zniekształceń.

1. Dla ciągów liter o ustalonej długości jako miarę zniekształcenia można przyjąć liczbę różniących się w obu ciągach liter.

2. Dla przebiegów ciągłych o określonym czasie trwania miarę odległości dwóch sygnałów $x_1(t), x_2(t)$ obserwowanych w przedziale $< -T, T >$ określimy jako

$$\varrho(x_1, x_2) = \max_{t \in < -T, T >} |x_1(t) - x_2(t)|. \quad (1)$$

Miara ta jest szczególnie dogodna, gdy przebiegi są obserwowane na ekranach lamp oscylograficznych.

3. Dla wiadomości takiego samego typu jak w przykładzie poprzednim może być dogodna miara

$$\varrho(x_1, x_2) = \sqrt{\int_{-T}^T [x_1(t) - x_2(t)]^2 dt}. \quad (2)$$

Pomimo, że w wielu przypadkach odbiega ona od naturalnej dla danego obiektu przeznaczenia wiadomości miary zniekształceń, bywa ona często stosowana ze względu na prostotę rachunkową.

Z kolei omówimy bardzo ważne pojęcie miary pojemności czyli objętości zbioru sygnałów.

Rozpatrzmy najprostszy przypadek, gdy zbiór używanych sygnałów jest skończony; jako przykład możemy wziąć alfabet, z którego budujemy wyrazy. Bardzo naturalną charakterystyką każdego podzbioru jest jego liczebność. Zwróćmy od razu uwagę na to, że wielkość ta ma własności miary zbioru w sensie matematycznym, to znaczy jest ona addytywną funkcją zbioru. Jeżeli będziemy rozpatrywali ciągi zbudowane z n elementarnych sygnałów omawianego typu, to zbiór tych ciągów jest znów zbiorem skończonym i podzbiory możemy charakteryzować za pomocą ich liczebności.

Z drugiej strony, ciągi zbudowane z n -elementów dogodnie jest przedstawiać jako punkty przestrzeni n -wymiarowej iloczynów Kartezjańskich. W takiej przestrzeni na ogół definiujemy miarę objętości jako iloczyn miar objętości określonych dla podprzestrzeni odpowiadających poszczególnym współrzędnym. Jest to uogólnienie typowego postępowania: powierzchnię prostokąta określamy jako iloczyn długości boków, objętość prostopadłościanu jako iloczyn długości krawędzi itd. Miara objętości danego podzbioru ciągów, określona w taki właśnie sposób, pokrywa się z miarą objętości zdefiniowaną poprzednio.

Przechodząc do zbiorów ciągłych, dla których z zasady wprowadzona jest miara odległości, zaczniemy od najprostszego przypadku, gdy sygnały można przedstawić jako punkty na prostej. Dla przykładu podamy impulsową modulację amplitudy, gdzie sygnałom są przyporządkowywane impulsy o tej samej szerokości, lecz o różnych wysokościach. Interesujące nas podzbiory będą z zasady odcinkami i jest rzeczą naturalną jako miarę objętości odcinka przyjąć jego długość. Będzie to zarazem największa odległość pomiędzy dwoma sygnałami należącymi do podzbioru, czyli

tak zwana średnica podzbioru. W podanym przykładzie modulacji impulsowej wielkość ta ma sens szczytowej amplitudy, a więc jest bardzo ważną charakterystyką techniczną.

Dla dodatkowego uzasadnienia podanej definicji objętości podzbiorów na prostej można się odwołać do rozpatrzonego poprzednio przypadku ziarnistego.

Czułość konkretnego obiektu odbierającego wiadomości ciągle jest z zasady skończona. Oznacza to, że dwie wiadomości bardzo mało odległe od siebie nie mogą być rozróżnione. Dolną granicę progu rozróżnialności wyznaczają efekty kwantowe i wahania związane z fluktuacjami termicznymi wewnątrz obiektu odbierającego wiadomości. W praktyce granica ta leży znacznie wyżej i określona jest na przykład przez tarcie (w przypadku urządzeń mechanicznych), nieostrość plamki (w przypadku lamp oscylograficznych) itd. Jeżeli czułość jest stała w całym zakresie, długość odcinka będzie proporcjonalna do liczebności zbioru rozróżnialnych sygnałów.

Dla zbiorów ciągów zbudowanych z elementarnych sygnałów jednowymiarowymch objętość określimy podobnie jak w przypadku ziarnistym, to jest jako miarę iloczynową w odpowiedniej przestrzeni wielowymiarowej.

Przejdźmy wreszcie do właściwego zagadnienia komunikacji.

Wskutek obecności przypadkowych zniekształceń po odbiorze sygnału y można podać jedynie prawo rozkładu prawdopodobieństwa (bądź gęstości prawdopodobieństwa) sygnałów nadawanych x — $p(x|y)$. Taka koncepcja jest konsekwentnie analizowana w pracy [13]. W praktyce wymaga się jednak, aby został określony jakiś sygnał *przypuszczalny*; będziemy go dalej oznaczać jako $x^*(y)$. Aby takie postępowanie miało w ogóle sens, trzeba aby zbiór sygnałów o dużym prawdopodobieństwie warunkowym był dostatecznie *skupiony* dookoła pewnej wartości. Bardzo ogólną miarę *skupienia* otrzymamy wprowadzając funkcję wagową $h(\varrho)$, gdzie $\varrho(x_1, x_2)$ jest funkcją pary sygnałów (w wielu przypadkach będzie ona miała bezpośrednio sens miary zniekształceń). Miarę *skupienia* dookoła $x^*(y)$ zdefiniujemy jako

$$E_{x|y} h[\varrho(x, x^*)], \quad (3)$$

gdzie

$E_{x|y}$ — oznacza uśrednienie na zbiór X sygnałów nadawanych x , gdy znany jest sygnał odebrany y .

Rozpatrzmy dwa przypadki, gdy zbiór sygnałów nadawanych X jest zbiorem, dla którego w sposób naturalny można określić miarę odległości $\varrho(x_1, x_2)$:

1. zakładając $h(\varrho) = \varrho^2$ otrzymujemy kryterium średniego rozrzutu kwadratowego;

2. zakładając $h(\varrho)=0$, gdy $\varrho < \delta$ i $h(\varrho)=1$, gdy $\varrho \geq \delta$; jako miarę rozrzutu otrzymujemy $P[\varrho(x, x^*) \geq \delta | y]$.

Przykład drugi nasuwa bardzo ważne uogólnienie. Rozrzut określony jest właściwie za pomocą dwóch liczb: promienia kuli δ dookoła sygnału x^* i prawdopodobieństwa zbioru możliwych sygnałów leżących poza tą kulą. Zamiast kuli możemy wprowadzić dowolny zbiór A_y i rozrzut charakteryzować za pomocą dwóch wielkości:

- 1) miary pojemności zbioru γA_y ,
- 2) miary prawdopodobieństwa $(x \in A_y | y)$.

Jak pokazano poprzednio, miara pojemności podzbiorów może być wprowadzona w sposób naturalny. Zwróćmy uwagę na to, że w tym ogólnym ujęciu nie podajemy właściwie przypuszczalnego sygnału, ale zamiast niego przypuszczalny zbiór A_y . Oczywiście, jeżeli obie miary są dostatecznie małe, jest to odpowiedź w pełni użyteczna.

Dotychczas rozpatrywano określony sygnał odebrany y . Dla pełnej charakterystyki systemu należy rozpatrzyć wszystkie możliwe sygnały y . Operując pojęciem wielkości średniej, wprowadzimy jako charakterystykę systemu wielkość

$$E_{Y \times | y} h[\varrho(x, x^*(y))]. \quad (4)$$

Zauważmy, że zamiast wielkości średniej można wprowadzić wartość

$$\max_{Y \times | y} E h[\varrho(x, x^*(y))]. \quad (5)$$

Podane poprzednio przykłady prowadzą do typowych charakterystyk systemów komunikacyjnych; na przykład w przypadku, gdy sygnały są realizacjami procesu stacjonarnego, a

$$\varrho(x_1, x_2) = \lim_{T \rightarrow \infty} \sqrt{\frac{1}{2T} \int_{-T}^T [x_1(t) - x_2(t)]^2 dt} \quad (6)$$

otrzymujemy kryterium Kołmogorowa-Wienera [12].

Po ustaleniu odpowiedniego kryterium dla charakterystyki danego systemu komunikacyjnego, wyłania się zagadnienie optymalnej metody wyboru przypuszczalnego sygnału $x^*(y)$ na podstawie sygnału odebranego y . Jest to zagadnienie bardzo istotne, silnie wiążące się z zasadniczymi problemami statystyki matematycznej. W pracy tej zajmować się nim nie będziemy.

3. MIARA NIEOKREŚLONOŚCI I MIARA ILOŚCI INFORMACJI

W pracy [10] zostało wprowadzone podstawowe pojęcie entropii i miary ilości informacji. Punktem wyjścia były rozważania dotyczące zbiorów ziarnistych. Podane dla nich definicje w drugiej części pracy zostały przeniesione mechanicznie na przypadek zbiorów ciągłych. Wydaje się, że na skutek tego uległy zatarciu ważne aspekty tych pojęć.

W pracy tej spróbujemy konsekwentnie zastosować wnioski wypływające z analizy komunikacji w obecności przypadkowych zniekształceń przeprowadzonej w rozdziale poprzednim. Podstawową koncepcją jest traktowanie entropii, którą w dalszym ciągu będziemy nazywali miarą nieokreśloności, jako specyficznej miary rozrzutu. Wyprowadzimy kilka twierdzeń, które będą to uzasadniały. Dalszą konsekwencją tego ujęcia jest próba powiązania miary nieokreśloności z innymi miarami rozrzutu.

Zajmiemy się w zasadzie zbiorami ciągłymi, a zbiory ziarniste, dla których zresztą pojęcie entropii i ilości informacji zostało dokładnie przeanalizowane, zwłaszcza w pracach [3] i [9], będziemy traktowali jako przypadek specjalny:

Przede wszystkim zwróćmy uwagę na to, że prawdopodobieństwo jest w zasadzie funkcją zbioru, a nie punktu. Z miarą prawdopodobieństwa zbioru $A - P(A)$ można związać funkcję punktu, a mianowicie gęstość prawdopodobieństwa $p(x)$. W tym celu trzeba wprowadzić drugą miarę, która ma sens miary objętości; będziemy ją oznaczali przez $-\gamma$. Zakładając, że obie miary spełniają pewne bardzo ogólne warunki ([2] rozdz. 39), dla dowolnego zbioru A mamy

$$P(A) = \int_A p(x) d\gamma, \quad (7)$$

co formalnie można napisać w postaci

$$p(x) = \frac{dP}{d\gamma}. \quad (8)$$

Jeżeli w przypadku ziarnistym za miarę objętości przyjąć liczebność zbioru, gęstość względem tej miary jest równa liczbowo prawdopodobieństwu.

Gdy dany sygnał można traktować jako ciąg sygnałów elementarnych, to dla ustalenia zależności pomiędzy własnościami statystycznymi całego ciągu a własnościami składników elementarnych, wydaje się naturalne wprowadzenie wielkości, która miałaby własności addytywne ze względu na dołączanie dodatkowych sygnałów elementarnych. Wielkość tę nazwiemy miarą nieokreśloności i będziemy oznaczali przez $H(x|y)$,
gdzie

x — jest sygnałem nadanym

y — odebrany.

Określimy ją przez aksjomaty:

I. $H(x|y)$ zależy tylko od gęstości prawdopodobieństwa $p(x|y)$,

II. jeżeli przez x_1, y_1 i x_2, y_2 oznaczmy dwie pary sygnałów nadanych i odebranych, przez (x_1, x_2) — sygnał złożony z pary sygnałów x_1, x_2 i jeżeli zachodzi

$$p[(x_1, x_2) | (y_1, y_2)] = p(x_1 | y_1) \cdot p(x_2 | y_2), \quad (9)$$

czyli sygnały x_1 i x_2 są warunkowo niezależne, to

$$H[(x_1, x_2) | (y_1, y_2)] = H(x_1 | y_1) + H(x_2 | y_2). \quad (10)$$

Zwróćmy od razu uwagę, że wielkość ta w istotny sposób zależy od miary objętości, w stosunku do której obliczamy gęstość.

Z prostych rozważań wynika, że jedyną (poza funkcjami bardzo nieregularnymi nie wchodzącymi w grę ze względów praktycznych) funkcją spełniającą aksjomaty jest funkcja

$$H(x | y) = -\log_b p(x). \quad (11)$$

Gdy miara γ jest wielkością bezwymiarową, wybór stałej jest sprawą jednostek. Ustalając $b=2$ przyjmujemy stosowaną w technice jednostkę „bit”. W dalszym ciągu będziemy pisali

$$H(x | y) = -\log p(x | y), \quad (12)$$

mając na myśli dowolną zasadę logarytmów, którą w razie potrzeby będziemy explicite zaznaczali.

Przejdźmy z kolei do pojęcia miary ilości informacji. Traktując możliwość zmiany prawa *a priori* $p(x)$ na prawo *a posteriori* $p(x | y)$ jako jedyną informację otrzymaną przez odbiór sygnału y , wprowadzić możemy pojęcie miary ilości informacji. Żądając, aby wielkość ta była addytywna ze względu na dołączanie dodatkowych elementów ciągu, naturalne będzie wprowadzenie następujących aksjomatów:

I. Miara ilości informacji o sygnale x uzyskana dzięki odbiorowi sygnału y — $I(x | y)$ zależy tylko od

$$p(x) \quad \text{ i } \quad p(x | y).$$

II. Gdy $p(x | y) = p(x)$ $I(x | y) = 0$.

III. Gdy zachodzi

$$p[(x_1, x_2)] = p(x_1) \cdot p(x_2), \quad (13)$$

$$\text{to} \quad p[(x_1, x_2) | (y_1, y_2)] = p(x_1 | y_1) \cdot p(x_2 | y_2), \quad (14)$$

$$I[(x_1, x_2) | (y_1, y_2)] = I(x_1 | y_1) + I(x_2 | y_2). \quad (15)$$

Podobna, aczkolwiek bardziej skomplikowana aksjomatyka podana była w pracy [13]. Podobne tu aksjomaty określają jednoznacznie miarę ilości informacji

$$I(x | y) = \log \frac{p(x | y)}{p(x)}. \quad (16)$$

Wielkość ta pozostaje w związku z miarą nieokreśloności odpowiadającym ich intuicyjnemu sensowi. Miara informacji równa jest różnicy miar nieokreśloności przed i po odbiorze sygnału.

Oznaczając przez $P_{X|Y}$ miarę prawdopodobieństwa warunkowego przy danym y , przez γ — dowolną miarę objętości, otrzymamy

$$\frac{dP_{X|Y}}{dP_X} = \frac{dP_{X|Y}/d\gamma}{dP_X/d\gamma} = \frac{p(x|y)}{p(x)}. \quad (17)$$

Z wyrażenia tego również widać, że miara ilości informacji nie zależy od wyboru miary objętości.

Miarę ilości informacji można traktować jako szczególny przypadek miary nieokreśloności. Objętość γ , względem której obliczamy gęstość prawdopodobieństwa warunkowego, jest w zasadzie dowolną miarą i można przyjąć, że jest to miara prawdopodobieństwa *a priori* P_X .

Z drugiej strony w rozdziale poprzednim pokazano, że gdy sygnały są ciągami sygnałów elementarnych, naturalne jest, aby miara objętości była określona jako miara iloczynowa. W przypadku przestrzeni nieskończenie-wielowymiarowych, których wprowadzenie jest nieuniknione, gdy się rozpatruje procesy stochastyczne, trzeba aby miara objętości całej przestrzeni była równa jedności. Można więc interpretować ją jako miarę prawdopodobieństwa, a miarę nieokreśloności jako specyficzną miarę ilości informacji.

Zajmijmy się średnią miarą nieokreśloności, którą oznaczmy jako

$$E_{X|Y} H(x|y) = H(X|Y). \quad (18)$$

Wielkość ta jest identyczna z wprowadzoną w pracy [10] entropią. Jak się należy spodziewać z aksjomatów, jest ona związana z pewnymi własnościami długich ciągów sygnałów. Związki te omówimy krótko w rozdziale 6.

Obecnie przystąpimy do przeanalizowania własności miary nieokreśloności ze względu na operacje dokonywane nad sygnałem odebrany i nadany.

4. WŁASNOŚCI ŚREDNIEJ MIARY NIEOKREŚLONOŚCI

Bardzo użyteczne w dalszym ciągu jest następujące twierdzenie [6]:

Twierdzenie 1

Niech P_1 i P_2 będą dwiema miarami prawdopodobieństwa, γ zaś miarą objętości, $p_1(x)$ i $p_2(x)$ niech będą gęstościami miar P_1 i P_2 względem miary γ . Mamy wówczas

$$\int_X [-\log p_1(x)] p_1(x) d\gamma \leq \int_X [-\log p_2(x)] p_1(x) d\gamma. \quad (19)$$

Znak równości zachodzi tylko wtedy, gdy prawie wszędzie względem γ

$$p_1(x) = p_2(x).$$

Sens tego twierdzenia jest następujący.

Jeżeli założyć jakieś prawo rozkładu i na podstawie dostatecznie dokładanych pomiarów (takich, aby częstości występowania były dostatecznie bliskie prawdopodobieństwom) obliczyć średnią miarę nieokreśloności, to wielkość minimalną uzyska się zakładając prawo prawdziwe.

Udowodnimy zasadnicze twierdzenie, które mierze nieokreśloności nadaje charakter wielkości mierzącej nieodwracalny wpływ przypadkowych zniekształceń, niedających się usunąć przez operacje ze strony urządzenia odbiorczego.

Twierdzenie 2.

Jeżeli sygnał odebrany zostaje poddany operacji T , to średnia miara nieokreśloności, liczona na podstawie sygnału przekształconego $y' = Ty$ jest niemniejsza niż obliczona na podstawie sygnału y :

$$\int_Y \int_{X|y} [-\log p(x|y)] dP_{X|y} dP_Y \leq \int_{Y'} \int_{X|y'} [-\log p'(x|y')] dP_{X|y'} dP_{Y'}. \quad (20)$$

Oznaczamy tu przez

- $P_{X|y}, P_{X|y'}$ — miary prawdopodobieństw warunkowych przy znanym y lub y' ,
 $p(x|y), p'(x|y')$ — gęstości prawdopodobieństw względem miary objętości γ ,
 $P_Y, P_{Y'}$ — miary prawdopodobieństw w przestrzeniach Y i Y' , przy czym $P_{Y'}$ określamy jako

$$P_{Y'} = T^{-1}P_Y.$$

Oznacza to, że miara zbioru B' sygnałów y' jest równa mierze zbioru B wszystkich sygnałów y , które są przez transformację T przekształcone w elementy zbioru B' .

D o w ó d

Średnią miarę nieokreśloności liczoną na podstawie znajomości sygnału y' zapiszmy w postaci

$$\int_{X \times Y'} [-\log p'(x|y')] dP_{X \times Y'}, \quad (21)$$

przy czym $P_{X \times Y'}$ jest miarą w przestrzeni punktów (x, y') zdefiniowaną wzorem

$$dP_{X \times Y'} = dP_{X|y'} \times dP_{Y'}. \quad (22)$$

Jeżeli analogicznie określimy miarę $P_{X \times Y}$, to zachodzi równanie

$$P_{X \times Y'} = T^{-1}P_{X \times Y}. \quad (23)$$

Na mocy twierdzenia o zmianie zmiennych ([2] — rozdz. 39)

$$\int_{\mathbf{x} \times \mathbf{y}'} [-\log p'(x|y')] dP_{\mathbf{x} \times \mathbf{y}'} = \int_{\mathbf{x} \times \mathbf{y}} [-\log p'(x|Ty)] dP_{\mathbf{x} \times \mathbf{y}}. \quad (24)$$

Zakończenie dowodu wynika z zależności (19).

Biorąc jako γ prawdopodobieństwo *a priori*, otrzymujemy analogicznie twierdzenie o zmniejszeniu miary ilości informacji.

Z kolei zajmujemy się wpływem operacji dokonywanych w urządzeniu nadawczym na miarę nieokreśloności *a priori*. Jest intuicyjnie zrozumiałe, że operacje te mogą ją w istotny sposób zmieniać, nawet jeżeli są to operacje jednoznacznie odwracalne. Zasadniczą rzeczą jest to, w jaki sposób zostaje zdefiniowana miara objętości przestrzeni sygnałów przekształconych. Rozpatrzmy typowe przykłady.

Jeżeli operacja polega po prostu na k -krotnym wzmocnieniu, czyli

$$x' = kx, \quad (25)$$

przy czym zdolność rozróżniania sygnałów Δx przez obiekt przeznaczenia wiadomości pozostaje bez zmian, to naturalne jest, że miarę objętości γ' w przestrzeni X' określimy wzorem

$$\gamma'(dx') = k\gamma(dx). \quad (26)$$

Postępowanie to można rozciągnąć na przypadek, gdy sygnały są ciągiem sygnałów elementarnych i każdy z wyrazów ciągu jest przekształcony według tej zasady. Do tego typu operacji można sprowadzić przesłanie sygnału przez filtr liniowy. Łatwo obliczyć, że zmiana miary nieokreśloności wynosi

$$\frac{1}{\Delta f} \int_{\Delta f} \log |\Phi(f)|^2 df, \quad (27)$$

gdzie

$\Phi(f)$ — jest funkcją przenoszenia filtru,

Δf — wstęgą zajętą przez sygnał.

Jako drugi przykład rozpatrzmy bardzo typową operację zwaną kwantyzacją.

Załóżmy, że sygnał x może być przedstawiony w postaci (x_1, x_2) . Niech $x' = Tx = x_1$; przez γ_1 i γ_2 oznaczmy miary objętości w przestrzeniach X_1 i X_2 . Jako miarę objętości w przestrzeni X przyjmujemy $\gamma = \gamma_1 \times \gamma_2$, jako zaś miarę objętości w przestrzeni X' przyjmujemy $\gamma' = \gamma_1$. Typową ilustracją może być przypadek zbiorów ziarnistych, gdzie wszystkie te miary są liczebnościami. Mamy więc

$$\begin{aligned} \int_{\mathbf{x} = \mathbf{x}_1 \times \mathbf{x}_2} [-\log p(x_1, x_2)] p(x_1, x_2) d\gamma_1 \times d\gamma_2 = \\ = \int_{\mathbf{x}_1} [-\log p(x_1)] p(x_1) d\gamma_1 + \int_{\mathbf{x}_1} p(x_1) d\gamma_1 \int_{\mathbf{x}} [-\log p(x_2|x_1)] p(x_2|x_1) d\gamma_2. \end{aligned} \quad (28)$$

Jeżeli drugi człon jest większy od zera, to

$$\int_x [-\log p(x)] p(x) d\gamma \geq \int_{x'} [-\log p'(x')] p'(x') d\gamma'. \quad (29)$$

Nierówność ta jest rozszerzeniem twierdzenia Shannona ([10] rozdz. 8) dla zbiorów ziarnistych.

Poprzednio podane przykłady operacji nie zmieniały w istotny sposób struktury sygnałów. W przykładzie pierwszym w zasadzie nie ulegała zmianie zajęta wstęga, w przykładzie drugim używane są takie same sygnały elementarne. Jako trzeci przykład podamy operację zmieniającą całkowicie strukturę sygnałów (na przykład modulacja częstotliwości). Można uznać, że w takim przypadku zdolność rozróżniania przez urządzenie odbiorcze sygnałów przekształconych zostaje zmieniona, natomiast nie zmienia się zdolność rozróżniania sygnałów odtworzonych, podawanych do obiektu przeznaczenia wiadomości. W takim przypadku naturalne jest zdefiniowanie pojęcia miary objętości w zbiorze sygnałów przekształconych X' jako

$$\gamma' = T^{-1}\gamma$$

(oznacza to, że objętość podzbioru A' sygnałów x' definiujemy jako objętość zbioru A wszystkich sygnałów x , które przez transformację T zostają przekształcone w elementy zbioru A'). Łatwo spostrzec, że założenia są analogiczne jak w twierdzeniu (20), a więc

$$\int_x [-\log p(x)] p(x) d\gamma \leq \int_{x'} [-\log p'(x')] p'(x') d\gamma'. \quad (30)$$

Warto zauważyć, że jest to nierówność przeciwna aniżeli nierówność (29), a więc przeciwna aniżeli klasyczna nierówność Shannona dla zbiorów ziarnistych.

5. ZWIĄZKI POMIĘDZY MIARĄ NIEOKREŚLONOŚCI A ŚREDNIM ROZRZUTEM KWADRATOWYM

Przystąpimy wpierw do określenia związku pomiędzy średnią miarą nieokreśloności a średnim rozrzutem kwadratowym.

Obliczymy minimum rozrzutu

$$\sigma^2 = \int_x \varrho^2[x, x^*(y')] p(x|y') d\gamma \quad (31)$$

przy pobocznych warunkach

$$\int_x [-\log p(x|y')] p(x|y') d\gamma = H(X|y') = \text{const}, \quad (32)$$

$$\int_x p(x|y') d\gamma = 1, \quad (33)$$

gdzie

x — jest sygnałem nadanym,

y' — przekształconym sygnałem odebranym.

$x^*(y')$ — przypuszczalnym sygnałem nadanym.

Proste obliczenie daje postać funkcji ekstremalizującej

$$p(x|y') = c_1 \exp[-c_2 \varrho^2(x, x^*(y'))]. \quad (34)$$

Z zależności (32) i (33) można wyznaczyć stałe c_1 i c_2 i określić wartość ekstremalną σ^2 jako funkcję H . Otrzymany wynik jest bardzo ogólny, gdyż miara nieokreśloności H jest liczona względem dowolnej miary objętości; w szczególności odnosi się on do miary ilości informacji.

Rozpatrzmy przykład, w którym stałe c_1 i c_2 dają się łatwo wyznaczyć. Niech sygnał x będzie punktem przestrzeni n — wymiarowej:

$$x = (x_1, x_2, \dots, x_n). \quad (35)$$

Miarę odległości określamy jako

$$\varrho^2(x', x'') = \sum_{ij} a_{ij} (x'_i - x''_i)(x'_j - x''_j), \quad (36)$$

przy czym, ażeby wielkość ta miała charakter odległości, będziemy żądali, aby macierz

$$\|a_{ij}\|$$

była dodatnio określona. Miarę objętości definiujemy wzorem

$$d\gamma = dx_1 \cdot dx_2 \cdot \dots \cdot dx_n. \quad (37)$$

Rozkład minimalizujący¹ ma postać

$$p(x|y') = c_1 \exp \left[-\frac{1}{2} c_2 \sum_{ij} a_{ij} (x_i - x_i^*)(x_j - x_j^*) \right]. \quad (38)$$

Z warunku (33) określamy

$$c_1 = \frac{|c_2 a_{ij}|^{\frac{1}{n}}}{(2\pi)^2}. \quad (39)$$

Obliczamy

$$H(X|y') = \log (2\pi e)^n |c_2 a_{ij}|^{-\frac{1}{2}}, \quad (40)$$

stąd wyznaczamy

$$\frac{1}{c_2} = \frac{1}{2\pi e} |a_{ij}|^{\frac{1}{n}} \exp \frac{2H(X|y')}{n}. \quad (41)$$

¹ Bardziej dokładnie rachunek wykazuje, iż występuje tu minimum.

Obliczamy wreszcie

$$\left\{ E_{x|y'} \varrho^2(x, x^*) \right\}_{\min} = \underbrace{\int \dots \int}_{\substack{\dots \\ n}} \sum_{ij} a_{ij} (x_i - x_i^*) (x_j - x_j^*) c_1 \exp \left[-\frac{1}{2} c_2 \sum_{ij} a_{ij} (x_i - x_i^*) (x_j - x_j^*) \right] dx_1 \dots dx_n = \frac{n}{2\pi e} |a_{ij}|^{\frac{1}{n}} \exp \frac{2H(X|y)}{n}. \quad (42)$$

Dla dowolnego rozkładu zatem

$$E_{x|y'} \varrho^2(x, x^*(y')) \geq \frac{n}{2\pi e} |a_{ij}|^{\frac{1}{n}} \exp \frac{2}{n} H(X|y') \quad (43)$$

po uśrednieniu na zbiór Y'

$$E_{x, Y'} \varrho^2[x, x^*(y')] \geq \frac{n}{2\pi e} |a_{ij}|^{\frac{1}{n}} E_{Y'} \exp \left[\frac{2}{n} H(X|y') \right], \quad (44)$$

ale

$$E_{Y'} \exp \frac{2H(X|y')}{n} \geq \exp E_{Y'} \frac{2H(X|y')}{n} \geq \exp \frac{2}{n} EY H(X|y), \quad (45)$$

przy czym nierówność pierwsza zachodzi, gdyż e^z jest funkcją wklęsłą, a nierówność druga wynika z monotoniczności tej funkcji i twierdzenia (20). Ostatecznie

$$\sum_{ij} \frac{a_{ij}}{|a_{ij}|^{\frac{1}{n}}} E_{x, Y'} [x_i - x_i^*(y')] [x_j - x_j^*(y')] \geq \frac{n}{2\pi e} \exp \frac{2}{n} EY H(X|y). \quad (46)$$

Twierdzenie to odgrywa rolę analogiczną jak twierdzenie Rao-Cramera ([4], rozdz. 10) z tym, że występują tu wartości średnie.

Szacowanie opiera się na znajomości jednej charakterystyki liczbowej — miary nieokreśloności. Jeżeli znane są dokładniejsze dane, można uzyskać szacowanie silniejsze. Załóżmy, że współrzędne x_i są niezależne od siebie i znane są ich miary nieokreśloności. Jako miarę odległości przyjmujemy

$$\varrho^2(x', x'') = \sum_i a_{ii} (x'_i - x''_i)^2, \quad (47)$$

przy czym

$$a_{ii} \geq 0.$$

Dla każdej ze współrzędnych będzie można zastosować nierówność (46)

$$E_{x_i, Y'} [x_i - x_i^*(y')]^2 \geq \frac{1}{2\pi e} \exp \frac{2EH(X_i|y)}{Y}. \quad (48)$$

Ostatecznie mamy

$$E_{x, Y'} \varrho^2[x, x^*(y')] \geq \frac{1}{2\pi e} \sum_i a_{ii} \exp \frac{2EH(X_i|y)}{Y}. \quad (49)$$

Nierówność (46) daje wartość granicy dolnej

$$\frac{n}{2\pi e} \exp \left[2 \sum_i \frac{1}{n} (EH(X_i | y) + \ln a_{ii}) \right], \quad (50)$$

gdź

$$EH(X | y) = \sum_i EH(X_i | y). \quad (51)$$

Ponieważ e^z jest funkcją wklęsłą, przeto

$$\begin{aligned} \frac{n}{2\pi e} \exp \left[2 \sum_i \frac{1}{n} (EH(X_i | y) + \ln a_{ii}) \right] &\leq \frac{n}{2\pi e} \sum_i \frac{1}{n} \exp [2EH(X_i | y) + \ln a_{ii}] = \\ &= \frac{1}{2\pi e} \sum_i a_{ii} \exp 2EH(X_i | y). \end{aligned} \quad (52)$$

Tak więc nierówność (49) otrzymana przy dodatkowych założeniach jest ostrzejsza, aniżeli nierówność (46).

Podamy wpierr obliczenie dolnej granicy błędu dla szczególnie prostego przypadku. Niech x_i oraz odpowiadające zakłócenie z_i mają rozkłady normalne o rozrzutach $\sigma_x(i)$ i $\sigma_z(i)$. Średnią miarę nieokreśloności łatwo obliczyć jako

$$H(X_i | y_i) = \frac{1}{2} \log 2\pi e \frac{\sigma_x^2(i) \cdot \sigma_z^2(i)}{\sigma_x^2(i) + \sigma_z^2(i)}, \quad (53)$$

gdzie $y_i = x_i + z_i$. Dolna granica błędu wynosi

$$\frac{1}{2\pi e} \exp 2H(X_i | y_i) = \frac{\sigma_x^2(i) \cdot \sigma_z^2(i)}{\sigma_x^2(i) + \sigma_z^2(i)}. \quad (54)$$

Rozpatrzmy z kolei następujące zagadnienie. Niech sygnał i addytywne zakłócenia mają strukturę realizacji procesów stacjonarnych $\mathbf{X}(t)$ i $\mathbf{Z}(t)$. Na podstawie odebranego sygnału $y(t) = x(t) + z(t)$ zostaje podany przypuszczalny sygnał $x^*(t)$. Jeżeli jako kryterium przyjąć wielkość średniego błędu kwadratowego

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T E_{\mathbf{XZ}} [x^*(t) - x(t - \Delta t)]^2 dt, \quad (55)$$

gdzie Δt jest opóźnieniem estymacji, i założyć, że operacja obliczania $x^*(t)$ na podstawie znajomości sygnału odebranego jest liniowa, otrzymuje się problematykę Kołmogorowa—Wienera (przykłady rozwiązania tego zagadnienia w pracach [5] i [12]). Spróbujemy ocenić czy możliwe jest zmniejszenie średniego błędu kwadratowego przez rozszerzenie klasy do-

puszczalnych operacji. Dopuszczymy klasę operacji obliczania przypuszczalnego sygnału $x^*(t)$ na podstawie odebranego $y(t)$ taką, że proces $\mathbf{X}^*(t)$ jest stacjonarny.

Dla określenia średniego błędu kwadratowego wystarcza w przypadku operacji liniowych znajomość funkcji korelacji, lub, co jest równoważne, widma mocy, natomiast do obliczenia miary nieokreśloności potrzebna jest pełna znajomość własności statystycznych procesu. Jeżeli znana jest tylko funkcja korelacji (jak to najczęściej ma miejsce w praktyce) naturalne jest wybranie spośród wszystkich procesów o tej samej funkcji korelacji, procesu, który jest przez funkcję jednoznacznie określony. Takim procesem jest proces normalny. Dlatego w dalszym ciągu ograniczymy się do procesów normalnych. Skorzystamy z przedstawienia spektralnego ([5] rozdz. 9).

$$\mathbf{X}(t) = \int_0^\infty \cos \omega t d\mathbf{U}_x^{(1)}(\omega) + \int_0^\infty \sin \omega t d\mathbf{U}_x^{(2)}(\omega), \quad (56)$$

gdzie $\mathbf{U}_x^{(1)}(\omega)$ i $\mathbf{U}_x^{(2)}(\omega)$ są procesami normalnymi nawzajem niekorelowanymi o niekorelowanych przyrostach. Zachodzi tu

$$E[d\mathbf{U}_x^{(1)}(\omega)]^2 = E[d\mathbf{U}_x^{(2)}(\omega)]^2 = \frac{1}{2\pi} d\bar{S}_x(\omega), \quad (57)$$

przy czym $\bar{S}_x(\omega)$ jest widmem mocy procesu. Dla uproszczenia będziemy zakładali, iż istnieje gęstość widmowa mocy $S_x(\omega)$, a zatem

$$d\bar{S}_x(\omega) = S_x(\omega) d\omega. \quad (58)$$

Proces $\mathbf{X}^*(t)$, estymator procesu $\mathbf{X}(t)$, jest według założenia procesem stacjonarnym, a więc analogicznie możemy go przedstawić w postaci

$$\mathbf{X}^*(t) = \int_0^\infty \cos \omega t d\mathbf{U}_{x^*}^{(1)}(\omega) + \int_0^\infty \sin \omega t d\mathbf{U}_{x^*}^{(2)}(\omega). \quad (59)$$

Procesy $\mathbf{U}_{x^*}^{(1)}$ i $\mathbf{U}_{x^*}^{(2)}$ możemy traktować jako estymatory procesów $\mathbf{U}_x^{(1)}$ i $\mathbf{U}_x^{(2)}$. Stosując wynik uzyskany dla poprzedniego przykładu — wzór (54) — mamy

$$E[d\mathbf{U}_x^{(1)} - d\mathbf{U}_{x^*}^{(1)}]^2 \geq \frac{S_x(\omega)S_2(\omega)}{S_x(\omega) + S_2(\omega)} \frac{d\omega}{2\pi}. \quad (60)$$

Ale

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T E[x^*(t) - x(t - \Delta t)]^2 dt = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T E \left\{ \int_0^\infty \cos \omega t [d\mathbf{U}_{x^*}^{(1)} - d\mathbf{U}_x^{(1)}] + \right.$$

$$\begin{aligned}
 + \int_0^{\infty} \sin \omega t [dU_{x^*}^{(2)} - dU_x^{(2)}]^2 dt &= \int_0^{\infty} [dU_{x^*}^{(1)} - dU_x^{(1)}]^2 + \int_0^{\infty} [dU_{x^*}^{(2)} - dU_x^{(2)}]^2 \geq \\
 &\geq \frac{1}{\pi} \int_0^{\infty} \frac{S_x(\omega) S_z(\omega)}{S_x(\omega) + S_z(\omega)} d\omega.
 \end{aligned} \quad (61)$$

Porównajmy to wyrażenie z wyrażeniem na minimalny błąd otrzymanym przez Wienera ([12] str. 88):

$$\{E[x^*(t) - x(t - \Delta t)]\}_{\min} = \frac{1}{\pi} \int_0^{\infty} \frac{S_x(\omega) \cdot S_z(\omega)}{S_x(\omega) + S_z(\omega)} d\omega + \Psi(\Delta t), \quad (62)$$

gdzie Δt jest opóźnieniem filtru, $\Psi(\Delta t) > 0$ oraz $\Psi(\Delta t) \rightarrow 0$, gdy $\Delta t \rightarrow 0$. Tak więc filtr optymalny z nieskończonym czasem opóźnienia daje minimum absolutne błędu kwadratowego i nic lepszego nie da się uzyskać za pomocą filtrów nieliniowych.

Otrzymany wynik można również interpretować w następujący sposób. Jak to zostało omówione w rozdziale 2, ściśle biorąc, znając sygnał odebrany y można podać jedynie prawo warunkowego rozkładu sygnałów nadanych. W przypadku filtru Wienera można uważać, że prawo to jest prawem normalnym z wartością średnią równą sygnałowi na wyjściu filtru i z rozrzutem równym średniemu błędowi kwadratowemu filtru. Zgodnie z twierdzeniem (20) taka zmiana faktycznego prawa rozkładu przez inne prawo jest związana ze zwiększeniem miary nieokreśloności lub stratą informacji. Z naszych rozważań wynika, że filtr optymalny z nieskończonym opóźnieniem takiej straty nie daje. Pociąga to za sobą ważną konsekwencję, którą zresztą można sprawdzić bezpośrednio. Stosując optymalną kwantyzację zbioru sygnałów nadawanych według metody podanej przez Shannona [10], można obliczyć średnią moc zbiorów sygnałów oryginalnych, którym zostaje przyporządkowany jeden i ten sam sygnał kwantyzowany. Nazywana ona bywa mocą szumu kwantyzacji i wyraża się przez (obliczenia pomijamy)

$$P_{S.kw} = \frac{1}{\pi} \frac{S_z(\omega) \cdot S_x(\omega)}{S_z(\omega) + S_x(\omega)}. \quad (63)$$

Jak widać ze wzoru (62) jest ona równa mocy błędu optymalnego filtru o nieskończonym opóźnieniu.

Wielkości te mają zupełnie inny sens. Pierwsza jest miarą minimalnego błędu wprowadzonego świadomie w urządzeniu nadawczym, aby potem

móc nieomal bezbłędnie przesyłać sygnał w obecności zakłóceń: druga wielkość jest miarą błędu kwadratowego, który popełniamy dokonując jak najskuteczniejszej operacji nad zniekształconym, nieskwantyzowanym sygnałem odebrany.

Na zakończenie tych rozważań zwróćmy uwagę na to, że rozumowanie nasze dotyczyło procesów normalnych. Gdy warunek ten nie jest spełniony, łatwo wykazać, że przy takiej samej mocy średniej procesu miara nieokreśloności sygnałów nadawanych jest mniejsza niż w przypadku normalnym. Zatem dolna granica błędu będzie mniejsza od granicy dla procesów normalnych. Istnieje więc możliwość uzyskania lepszych wyników przez zastosowanie filtrów nieliniowych. Konieczne będą jednak dokładniejsze dane statystyczne o procesie, aniżeli sama funkcja korelacji.

6. ZWIĄZKI POMIĘDZY MIARĄ NIEOKREŚLONOŚCI A PRAWDOPODOBIENSTWEM BŁĘDÓW

Zajmiemy się obecnie związkiem pomiędzy miarą nieokreśloności i miarą rozrzutu, określoną jako wielkość prawdopodobieństwa P_{Bl} tego, że sygnał nadany x leży poza zbiorem A o objętości — γA . Zwróćmy uwagę na to, że definicja ta obejmuje zarówno przypadek ciągły jak i ziarnisty. W przypadku ogólnym prosty związek zachodzi dla długich ciągów sygnałów, przy założeniu, że $P_{Bl} \rightarrow 0$. Jest to w zasadzie znane twierdzenie Shannona ([11] rozdz. 21).

Twierdzenie

Weźmy pod uwagę przestrzeń ciągów zbudowanych z n elementarnych niezależnych sygnałów X_n . Niech w przestrzeni X_1 wprowadzona będzie miara objętości γ_1 . Miarę objętości w X_n zdefiniujemy jako $[\gamma_1]^n$. Dla dowolnego $\varepsilon > 0$ i $\delta > 0$ można znaleźć takie N_0 oraz dla każdego n taki zbiór A_n , że dla $n > N_0$ będzie

$$\left| \frac{\log \gamma_n A_n}{n} - H(X_1) \right| < \delta \quad \text{oraz} \quad P_{Bl} = P(X_n - A_n) < \varepsilon. \quad (64)$$

Twierdzenie to może być rozciągnięte na ciągi o członach zależnych, które obowiązuje prawo liczb wielkich, tak zwane ciągi ergodyczne. Jego znaczenie praktyczne nie jest wielkie ze względu na podwójnie asymptotyczny charakter zależności ($n \rightarrow \infty$, $P_{Bl} \rightarrow 0$). Również mało użyteczna jest jego dalsza formalna konsekwencja — twierdzenie o pojemności kanału [1].

Ustalimy z kolei związek pomiędzy wielkościami H , P_{Bl} , γA dla sygnału elementarnego. Załóżmy, że rozkład prawdopodobieństwa $p(x)$ jest taki, że na zewnątrz zbioru X_0 o mierze objętości skończonej mamy $p(x) = 0$.

Stosując metodę rachunku wariacyjnego, jak dla szacowania minimalnego błędu kwadratowego, otrzymujemy

$$H = \int_A [-\log p(x)] p(x) d\gamma + \int_{X_0-A} [-\log p(x)] p(x) d\gamma \leqslant (1 - P_{Bt}) \left(-\log \frac{1 - P_{Bt}}{\gamma A} \right) + P_{Bt} \log \frac{P_{Bt}}{\gamma (X_0 - A)}. \quad (65)$$

Proste przekształcenie daje

$$H_{\max} - H \geqslant (1 - P_{Bt}) \log \frac{\gamma X_0}{\gamma A} + P_{Bt} \log \frac{\gamma X_0}{\gamma (X_0 - A)} - H_0(P_{Bt}), \quad (66)$$

przy czym

$$H_{\max} = \log \gamma X_0 \quad (67)$$

jest maksymalną miarą nieokreśloności, która zostaje osiągnięta wówczas, gdy rozkład prawdopodobieństwa wewnątrz zbioru X_0 jest równomierny.

$$H_0(P_{Bt}) = -P_{Bt} \log P_{Bt} - (1 - P_{Bt}) \log (1 - P_{Bt}) \quad (68)$$

jest miarą nieokreśloności, gdy wiadomo, że są dwie możliwości z prawdopodobieństwami P_{Bt} i $1 - P_{Bt}$.

Z wyrażenia (66) można uzyskać szacowanie maksymalnego prawdopodobieństwa błędu, to jest jego najbardziej niekorzystnej wielkości, gdy

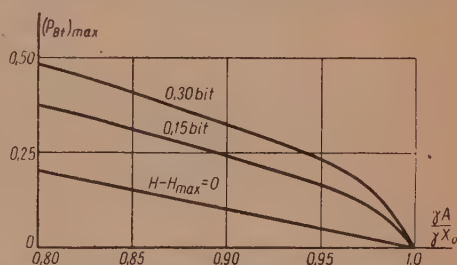
$$\frac{\gamma A}{\gamma X_0} \rightarrow 1. \quad (69)$$

Wyniki obliczeń przedstawiono na rys. 2. Podano tam tylko małe wartości parametru $H_{\max} - H$, co odpowiada rozkładowi zbliżonym do równomiernego; dla rozkładów znacznie różniących się od niego, szacowanie traci sens. Na podstawie krzywych z rys. 2 można ocenić sprawność estymacji w konkretnych przypadkach.

Zwróćmy wreszcie uwagę na to, że sposób podawania przypuszczalnego sygnału jest równoważny zastąpieniu faktycznego prawa rozkładu, prawem rozkładu określonym wzorami:

$$\left. \begin{aligned} p(x) &= \frac{1 - P_{Bt}}{\gamma A} & \text{gdy } x \in A, \\ p(x) &= \frac{P_{Bt}}{\gamma (X_0 - A)} & \text{gdy } x \in X_0 - A. \end{aligned} \right\} \quad (70)$$

W oparciu o twierdzenie (20) można określić sprawność przekazywania informacji.



Rys. 2. Wykres zależności maksymalnego prawdopodobieństwa błędu od stosunku $\gamma A / \gamma X_0$.

Jako ostatni przykład rozpatrzmy system, gdzie używane są dwa sygnały, na przykład alarm i brak alarmu. Metoda podawania przypuszczalnego sygnału może być scharakteryzowana przez dwa typy błędów

1. Prawdopodobieństwo $P(x_2^*|x_1)$ tego, że gdy nadany jest sygnał x_1 jako przypuszczalny $x^*(y)$ podany będzie przez urządzenie odbiorcze sygnał x_2 .

2. Prawdopodobieństwo $P(x_1^*|x_2)$ tego, że gdy nadany jest sygnał x_2 jako przypuszczalny $x^*(y)$ podany będzie przez urządzenie odbiorcze sygnał x_1 .

Całkowite prawdopodobieństwo błędu wynosi

$$P_{Bt} = P(x_1) P(x_2^*|x_1) + P(x_2) P(x_1^*|x_2), \quad (71)$$

gdzie $P(x_1)$ i $P(x_2)$ są prawdopodobieństwami *a priori*.

Prosta metoda prowadzi do zasady podawania przypuszczalnego sygnału y minimalizującego ten błąd; jako przypuszczalny sygnał $x^*(y)$ podawany jest sygnał x_1 lub sygnał x_2 w zależności od tego czy liczba Δ określona wzorem

$$\Delta = \frac{P(x_1) p(y|x_1)}{P(x_2) p(y|x_2)} \quad (72)$$

jest większa czy mniejsza od jedności ([7] rozdz. 7).

Zauważmy, że podanie średniego prawdopodobieństwa błędu jest równoważne zastąpieniu prawa rozkładu

$$P(x \neq x^*(y)|y) \quad (73)$$

przez prawo uśrednione

$$P_{Bt} = E_{x^*} E_{y|y \in Y_i} P(x \neq x^*(y)|y), \quad (74)$$

gdzie

$E_{y, y \in Y_i}$ — oznacza uśrednienie przy warunku, że y należy do zbioru Y_i , w którym jako przypuszczalny podaje się sygnał x_i^* ,

E_{x^*} — oznacza uśrednienie na możliwe przypuszczalne sygnały x^* .

Obliczymy zwiększenie miary nieokreśloności w tym przypadku.

Dla uproszczenia przyjęto, że wielkość przekazywana jest jednowymiarowa. Zakładamy, że zakłócenia mają rozkład normalny o rozrzucie σ . Przez Δ oznaczamy odległość sygnałów. Miara nieokreśloności, po podaniu sygnału przypuszczalnego według wyżej opisanej metody, wynosi

$$EH'(X|x^*) = -P_{Bt}^{\min} \log P_{Bt}^{\min} - (1 - P_{Bt}^{\min}) \log (1 - P_{Bt}^{\min}). \quad (75)$$

Średnia miara nieokreśloności, gdy znany jest sygnał odebrany y , przy uczynionych założeniach może być obliczona ze wzoru

$$EH(X|y) \approx 3,05 - H(Y). \quad (76)$$

Graficznie obliczono wielkość $H(Y)$. Wyniki pokazane są na rys. 3, z którego widać, że sprawność systemu jest duża. Niewiele tracimy podając przypuszczalny sygnał na podstawie pojedynczego odebranego sygnału. Gdyby podawać przypuszczalny ciąg na podstawie odebranego ciągu dostatecznie długiego, otrzymalibyśmy miarę nieokreśloności zbliżoną do granicznej.

Charakterystyka systemu za pomocą jednej liczby, całkowitego prawdopodobieństwa błędu P_{Bt} , jest mało dokładna. W systemach bardziej precyzyjnych można uważać, że znane są prawdopodobieństwa błędów pierwszego i drugiego rodzaju (najczęściej jeden z nich narzucony jest z góry, a drugi minimalizujemy przez wybór odpowiedniej zasady podawania przekształconego sygnału — (na przykład [4] rozdział 13).

Pominiemy, podobny jak poprzednio, problem obliczenia straty informacji związanej z taką metodą podawania przypuszczalnego sygnału, a zatrzymamy się przy zagadnieniu wpływu poziomu błędów pierwszego i drugiego rodzaju na ilość ostatecznie przekazywanej informacji.

Dla uproszczenia zapisów wprowadzimy oznaczenia

$$P(x_1)=a \quad P(x_2)=b \quad P(x_2^*|x_1)=a \quad P(x_1^*|x_2)=\beta \quad (77)$$

oraz

$$\Phi(x) = -x \log x - (1-x) \log (1-x). \quad (78)$$

Średni zysk informacji wynosi

$$I = \Phi[a(1-a) + \beta b] - a\Phi(a) - b\Phi(\beta). \quad (79)$$

Nie wnikając w żmudną analizę tej funkcji, zbadamy związek pomiędzy zyskiem informacji a błędem całkowitym

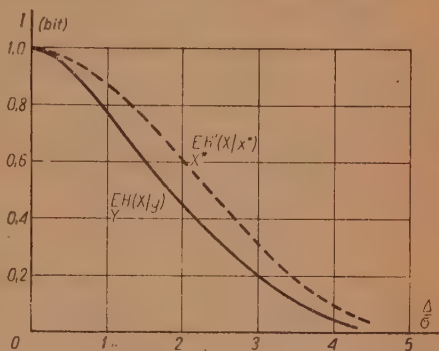
$$P_{Bt} = a + b\beta. \quad (80)$$

Dla uproszczenia założmy $a=b=0,5$, otrzymujemy wówczas

$$I = \Phi\left(\frac{1}{2} - P_{Bt} + \beta\right) - \frac{1}{2} \Phi(2P_{Bt} - \beta) - \frac{1}{2} \Phi(\beta). \quad (81)$$

Założmy $x = P_{Bt} - \beta$; wyrażenie (81) przyjmie wtedy postać

$$I = \Phi\left(\frac{1}{2} - x\right) - \frac{1}{2} \Phi(P_{Bt} + x) - \frac{1}{2} \Phi(P_{Bt} - x). \quad (82)$$



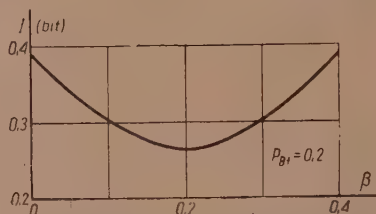
Rys. 3. Wykresy miar nieokreśloności dla systemu idealnego i systemu minimalizującego średnie prawdopodobieństwo błędu w zależności od stosunku sygnału do szumu.

Łatwo spostrzec, że wielkość ta osiąga minimum dla $x = \frac{1}{2}$. Wykres I jako funkcji β przy ustalonym P_{Bt} pokazano na rys. 4, wykresy $I_{\max}$ i $I_{\min}$ jako funkcje P_{Bt} podano na rys. 5.

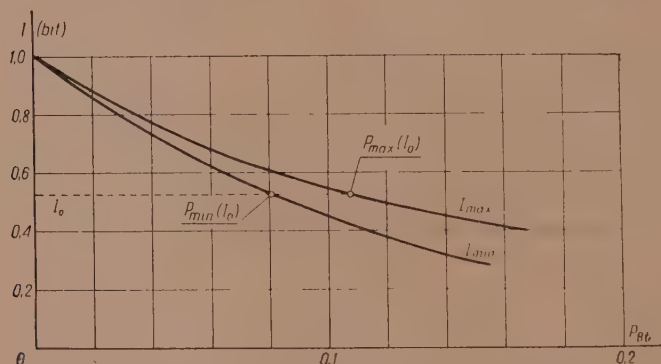
Uzyskane zależności ustalają bardzo istotny związek pomiędzy prawdopodobieństwem całkowitym błędu a zyskiem informacji. Urządzenie odbiorcze musi osiągnąć pewną ilość informacji I_0 . Błąd całkowity nie może być większy od $P_{\max}(I_0)$ (rys. 5); α i β mogą się zmieniać, lecz tak, aby

$$\alpha + \beta \leq 2P_{\max}(I_0). \quad (83)$$

Nierówność ta daje inne oświetlenie tak zwanej zasady komplementarności błędów, omawianej w pracy [8].



Rys. 4. Wykaz miary nieokreśloności w zależności od błędu drugiego rodzaju.



Rys. 5. Wykresy minimalnej i maksymalnej miary nieokreśloności w zależności od całkowitego prawdopodobieństwa błędu.

WYKAZ LITERATURY

1. Bloch E. L., Charkiewicz A. A.: *Gieometricheskaja teorija poroga propusknnoj sposobnosti sistemy swiazi*. Radiotekhnika nr 7, 1955.
2. Chałmosz: *Teorija miery*, Moskwa 1953.
3. Chinczyn A.: *Poniatije entropii w teoriji wierojatnostiej*. U. M. N. nr 3, 1953.
4. Fisz M.: *Rachunek prawdopodobieństwa i statystyka matematyczna*. Warszawa 1954.
5. Jagłom A.: *Wwiedienije w teoriju stacionarnych sluczajnych funkcij*. U. M. N. nr 5, 1952.
6. Kullback S.: *On information and sufficiency*. An. math. Stat. 1952, s. 79.
7. Lawson J.: *Threshold signals*. N. York 1950.
8. Loeb J.: *Les deux types d'erreurs*. Ann. Tell. II, 1954.

9. Mc Millan B.: *The basic theorems of information theory*. An. Math. Stat. 1953, s. 196.
10. Shannon C.: *The mathematical theory of communication*. Urbana 1949.
11. Shannon C.: *Communication in presence of noise*. P. I. R. E. I, 1949.
12. Wiener N.: *Extrapolation, interpolation and smoothing of stationary time series*. J. Wiley, 1949.
13. Woodward P., Davies I: Proc. I. E. E. pt. III 1952. 99. 45.

Г. СЕЙДЛЕР

О ПОНЯТИИ МЕРЫ НЕОПРЕДЕЛЁННОСТИ (ЭНТРОПИИ)

Резюме

1. *Введение.* Понятие энтропии имеет простой смысл для множеств дискретных сигналов.

Если определяющие сигнал параметры могут изменяться непрерывно, то энтропия *априорная* и *опостериорная*, рассчитанные формально подобным образом как для дискретных распределений позволяют получить асимптотическое выражение для числа сигналов, могущих быть безошибочно переданными. Чтобы приблизиться к таким условиям передачи необходимо провести кодирование. Таким образом, однако, непрерывная система в действительности замещается дискретной. Между тем на практике очень часто применяются не квантизированные системы, к тому же не предвиденные к квантизации. Возникает вопрос, можно ли энтропии приписать иную интерпретацию, чем величины описывающей потенциальные возможности системы после кодирования и следовательно пользоваться ею как характеристикой систем фактически непрерывных?

Обращено также внимание на то, что на практике нет возможности ни нужды уменьшить вероятность ошибок до нуля. Теперь возникает второй вопрос, может ли энтропия быть полезной в качестве характеристики системы (применяющей непрерывное или дискретное множество сигналов), в которой допустима небольшая, но конечная вероятность ошибок?

2. *Числовые характеристики систем связи.* Если сигналы могут изменяться непрерывным образом, то свойства объекта, к которому направлены сообщения, по большей части позволяют определить числовую меру расстояния сигналов (меру искажения). Вследствие наличия случайных искажений (как напр. наложения шума) после принятия сигнала y , точно говоря, можно привести только закон условного распределения передаваемых сигналов x . Но на практике требуется передача какого-нибудь предполагаемого сигнала. Обозначим его $x^*(y)$. Как числовую характеристику качества системы примем величину, определённую формулой (4).

Обозначение: величина $\varrho(x, x^*)$ обозначает расстояние переданного сигнала x от сигнала предполагаемого x^* ; $h(\varrho)$ — соответствующую весовую функцию, напр. $h(\varrho) = \varrho^2$ или $h(\varrho) = 0$, когда $\varrho \leq \delta$, а также $h(\varrho) = 1$; когда $\varrho > \delta$ $\overline{E E}$ — операцию усреднения распространённую на множество всех передаваемых и принимаемых сигналов.

Конечная чувствительность объекта, к которому направлены сообщения, устанавливает порог отличаемости параметров определяющих сигнал. Это во многих случаях ведёт к определению понятия меры объёма подмножества передаваемых сигналов. В таком случае более соответственной является подача

вместо предполагаемого сигнала — предполагаемое множество A_y сигналов x и характеристика системы при помощи двух чисел:

1. средней вероятности, что переданный сигнал находится вне множества A_y
2. среднего объёма множества A_y . Эти два числа, также как и величину, определённую формулой (4) можно в общем назвать мерами разброса.

3. Основные свойства энтропии непрерывных множеств. При анализе понятия энтропии влияние преобразований передаваемых и принимаемых сигналов имеет основное значение. В случае непрерывных распределений энтропия зависит существенным образом от меры объёма, по отношению к которой рассчитываются плотности вероятности. В особенности, если в качестве объёма принять вероятность *априорную* — то получается мера количества информации; поэтому мы не будем разделять эти понятия.

Если принять во внимание преобразование принимаемых сигналов, то кажется, что следующая теорема имеет основное значение: никакие операции по отношению к принятому сигналу не могут уменьшить средней энтропии. Формула (20).

Обозначения: $p(x|y)$ и $p'(x|y')$ являются условными плотностями вероятности, когда известен первоначальный принятый сигнал y или преобразованный сигнал y' ; плотности считаются по отношению к одной и той же мере объёма.

Преобразования передаваемых сигналов в противоположность к дискретному распределению могут как уменьшать, так и увеличивать меру неопределённости. Для преобразований типа квантизации имеет место неравенство (29).

(Обозначения $p(x)$ и $p'(x')$ являются плотностями вероятности первоначального сигнала x и преобразованного сигнала x').

Если меру объёма в пространстве преобразованных сигналов x' мы определим как $y' = T^{-1}y$ что является удобным определением напр. для частотной модуляции, то имеет место неравенство (30) т. е. неравенство обратное неравенству (29), которое в свою очередь является обобщением теоремы Шэннона для дискретных распределений.

4. Зависимость между энтропией и другими мерами разброса. Основным понятием является принятие энтропии в качестве меры разброса. Приведенные в предыдущем параграфе свойства позволяют оценить при помощи энтропии нижние пределы других мер разброса. Это свойство придаёт энтропии более общее значение. Для квадратичного разброса имеет место зависимость (46), где:

a_{ij} — элементы положительно определённой матрицы, x_i $i=1,2,\dots,n$ координаты определяющие сигнал $x=(x_1,\dots,x_n)$

$H(X/y)$ — относительная мера неопределённости, причём плотности вероятности считаются по отношению к мере объёма $dy=dx_1 \cdot dx_2 \dots dx_n$

$x_1^*(y')$ — предполагаемое значение параметра x_1 , поданное на основании известного y'

Простое применение этого неравенства ведёт к следующим заключениям. Когда сигналы и помехи имеют структуру стационарных гауссовых процессов, то средняя мощность погрешности оптимального фильтра Винера с бесконечно большим запаздыванием равна мощности шума квантизации при оптимальном кодировании по Шэннону. При помощи линейного фильтра получается абсолютный минимум квадратичной погрешности.

Вторая группа зависимостей — это зависимости между энтропией и мерой разброса, определённой как средний объём предполагаемого множества и вероятность ошибки.

Если вне множества X_0 $p(x)=0$, имеет место неравенство (66), где P_{Bl} вероятность, что сигнал находится вне множества A , γ — мера объема, а величины $H_{\max}$, $H_0(P_{Bl})$ определены формулами:

$$H_{\max} = \log \gamma x_0, H_0(P_{Bl}) = -P_{Bl} \log P_{Bl} - (1 - P_{Bl}) \log (1 - P_{Bl}).$$

Эта зависимость позволяет нам оценить P_{Bl} и γA . Получены также оценки для погрешностей первого и второго рода в случае систем применяющих два сигнала (напр. тревожный сигнал и отсутствие тревожного сигнала) оптимизированных по методу Неймана—Пирсона и Зигерта. Результаты представлены на диаграммах (4) и (5).

J. SEIDLER

ON THE CONCEPT OF ENTROPY

Summary

1. *Introduction.* The concept of entropy has a simple meaning for discrete sets of signals. If parameters determining the signal can vary continuously, the a priori and a posteriori entropies, calculated in a formally similar way as it is in discrete cases, allow to obtain the asymptotic expression for maximal quantity of signals that can be transmitted without error. The coding of signals is necessary in order to approach such conditions of transmitting. However a continuous system is thus actually replaced by a discrete system. But in practice very often such systems are in use, that are neither quantized nor meant to be quantized. A question arises, whether the entropy can be interpreted otherwise than as a quantity describing potential possibilities of a system when coded, and consequently, whether it can be used as a characteristic of the actually continuous systems.

Let us notice, too, that in practice it is neither possible nor necessary to reduce the probability of the error to zero. And thus, the next question arises: can entropy be useful as a characteristic of a system, continuous or discrete, permitting a small but finite probability of error.

2. *Numerical characteristics of communication systems.* If signals can vary continuously, the properties of the object the messages are directed to, lead to the determination of a measure of distance between signals (measure of distortion). When a signal — y is received, then — owing to some random distortions present (noise etc) — strictly speaking, only a law of conditional distribution of transmitted signals x can be stated. But in practice, some expected signal is required. We shall denote it by $x^*(y)$.

As a numerical characteristic of the system, we accept the quantity determined by formula (4).

(Notations: $q(x, x^*)$ denotes the distance between the transmitted signal x and expected signal x^* ,

$h(q)$ — weight function e. g. $h(q)=q^2$ or $h(q)=0$, if $q \leq \delta$ and $h(q)=1$ if $q > \delta$, and

E — the averaging operation over the set of all received and transmitted signals).

The existence of the threshold of ability to distinguish two messages, is the feature of any concrete object receiving messages. This feature permits in many cases to define the measure of volume of subsets of signals. Then it is more conse-

quent to give an expected subset Ay of signals x , instead of the expected signal. The system can be characterized by two numbers:

1. The average probability that the transmitted signal lies outside the subset Ay .
2. The average volume of the subset Ay .

These two quantities and the one defined by formula (4), may be called the measure of spreading of the signals.

3. *The fundamental properties of entropy of continuous distributions.* When analyzing the concept of entropy, it seems that the influence of transformations performed on the received and transmitted signals is essential.

It should be kept in mind, that in the continuous case the entropy depends in an essential way on the measure of volume in respect to which the probability density is calculated. Particularly, if the a priori probability is taken as the measure of volume, the quantity of information is obtained. Therefore, these two concepts will not be distinguished here.

Let us take at first the operations on received signals. The following theorem seems to be fundamental:

No operation on received signals can decrease the conditional entropy. See formula (20).

Notations: $p(x|y)$ and $p'(x|y')$ denote the conditional probability distributions when the original received signal y , or the transformed y' is known. Both densities are calculated in respect to the same measure of volume.

Contrary to the discrete case, the transformations of transmitted signals can either increase or decrease the entropy. Inequality (29) holds for the transformations of the quantization type.

(*Notations:* $p(x)$ and $p'(x')$ denote the probability densities of the original signal x and of the transformed x')

If the measure of volume of signals transformed in the space x' is defined as $y' = T^{-1}y$, which is for example appropriate for frequency modulation, then inequality (30) holds.

This is the reversed inequality (29), which in turn presents the generalization of Shannon's classical theorem for the discrete case.

4. *Relations between entropy and other measures of spreading.* The fundamental idea of this paper is the interpretation of the entropy as a measure of spreading. The properties discussed in the preceding paragraph permit to estimate boundaries of other measures of spread by means of entropy. Thus, a wider significance is given to entropy. For dispersion, a relation given by formula (46) can be stated.

(*Notations:* a_{ij} denotes elements of a positively defined matrix,

x_i $i=1, 2, \dots, n$ — the coordinates determining the signal $x=(x_1, \dots, x_n)$,

$H(X|y)$ — the conditional entropy with probability density calculated in respect to the measure of volume defined as $dy = dx_1 \cdot dx_2 \dots dx_n$

$x_i^*(y')$ — the expected value of parameter x_i calculated when the transformed received signal y' is known).

An application of this inequality leads to the following conclusions. Let us suppose that signals and noise can be represented by stationary gaussian processes. Then, the average error power in the infinite delay Wiener-filter and the power of quantization noise — when optimal coding is applied — are equal. It follows also that the linear filter gives in this case the absolute minimum of error, and nothing better can be achieved by nonlinear filters.

Now, let us consider the relation between entropy and the measure of spreading defined as the average volume of the expected subset of signals and the error probability P_{Bl} .

If outside the set X_0 $p(x)=0$, formula 66 holds,
(Notations: P_{Bl} denotes the probability that x lies outside A ,
 γ — the measure of the volume)

$$- \text{ then } H_{\max} = \log \gamma x_0, \quad H_0(P_{Bl}) = -P_{Bl} \log P_{Bl} - (1 - P_{Bl}) \log (1 - P_{Bl}).$$

This formula permits the simultaneous estimation of P_{Bl} and γA .

For the systems using two signals (e.g. alarm and rest) optimized in the sense of Neyman-Pearson and Siegert, we have also obtained formulae to estimate the errors of the first and second kind. See Fig. 4 and Fig. 5.

ERRATA

Zeszyt 1 — Tom V — 1956

Str.	Wiersz		Jest	Powinno być
	od dołu	od góry		
88	5		$\Delta''_E = \frac{r_1}{4\pi} \sqrt{\frac{\mu}{\varepsilon}} \int \varphi (\mathbf{E} \cdot d\mathbf{l}),$	$\Delta''_E = \frac{r_1}{4\pi} \sqrt{\frac{\mu}{\varepsilon}} \int \varphi (\mathbf{H} \cdot d\mathbf{l}),$
89		9	są w strefie	nie mają w strefie
93	1		$\gamma'_E =$	$\gamma'_H =$
103		13	$= \text{const} \frac{\partial E_o}{\partial n} = 0$	$\mathbf{E} = \text{const} \frac{\partial E_n}{\partial n} = 0$
106	1		$= 0.$	$\Delta''_H = 0.$
144	6		$\varphi = \sqrt{\quad}$	$\varphi \approx \sqrt{\quad}$
147	4		equation	equations
147	3		(27) and (28)	(27') and (28')
147	3		from	form
157	3		$A'_r = \mathbf{A}'_\varphi = 0$	$A'_r = A_\varphi = 0$
167		6	$\Phi' = J_o \left(\frac{P_n}{R} r \right) \left[\dots \right]$	$\Phi' = I_o \left(\frac{p_n}{R} r \right) \left[\dots \right]$
167	7		$e^{-j\gamma_1 l} - e^{-j\gamma_2 l} =$	$e^{-j\gamma_1 l} - e^{j\gamma_2 l} =$
176	14		$\varepsilon_{wz2} = 1 - \frac{b^2}{(\omega^2 + \gamma_2^2 v)^2} =$	$\varepsilon_{wz2} = 1 - \frac{b^2}{(\omega^2 + \gamma_2^2 v_0)^2} =$
186	1		$= \Phi_- + j\Phi_-$	$= \Phi_- + j\Phi_-$
191		2	$\int_0^R r I_o^2 \left(\frac{p_n}{R} r \right) dr =$	$\int_0^R r I^2 \left(\frac{p_n}{R} r \right) dr =$
202		11	$\Phi \mathbf{E} d\sigma =$	$\oint \mathbf{E} d\bar{\sigma} =$
202		13	$\int \mathbf{E} d\sigma$	$\int \mathbf{E} d\bar{\sigma} =$
202	13		$\Phi \mathbf{E} d\sigma =$	$\oint \mathbf{E} d\bar{\sigma} =$

WYTYCZNE DLA AUTORÓW

Komitet Redakcyjny prosi autorów o ułatwienie prac redakcyjnych związanych z przygotowaniem do druku nadesłanych artykułów przez przestrzeganie podanych wytycznych przy przygotowaniu maszynopisu:

1. Prace powinny być napisane pismem maszynowym, na pojedynczych arkuszach formatu A4, jednostronnie, z interlinią (co drugi wiersz), z marginesem 3 cm z lewej strony. Stronice numerowane. Artykuły należy nadsyłać w dwóch egzemplarzach.
2. Wzory i oznaczenia należy wpisywać ręcznie, czytelnie, używając jedynie liter łacińskich i greckich. Wskaźniki niżej liter i wykładniki potęg pisać należy szczególnie dokładnie i wyraźnie.
3. Każda praca powinna być zaopatrzona w krótkie streszczenie (analizę) w języku polskim do 25 wierszy maszynopisu oraz obszerniejsze streszczenia (do 20% objętości artykułu) w języku rosyjskim oraz angielskim, francuskim lub niemieckim. W razie niemożności nadesłania streszczenia w języku obcym autor dostarcza odpowiednie streszczenie w języku polskim w trzech egz. z jednoczesnym wpisaniem, o ile możliwości, terminologii w języku rosyjskim oraz w innym języku obcym.
4. Rysunki, wykresy i fotografie należy wykonywać na oddzielnych arkuszach z podaniem kolejnych numerów rysunków. W tekście i na marginesie, obok właściwego tekstu, należy podać jedynie odpowiadający numer rysunku. Ostateczne wykonanie rysunków obowiązuje Redakcję.
5. Wszystkie rysunki, wykresy i fotografie należy nazywać w tekście rysunkami (skrót: rys.) i nie używać określeń jak figura, szkic, fotografia. U samego dołu rysunku (a przy fotografiach — na odwrocie) należy wpisać czytelnie numer rysunku, napis pod rysunkiem, tytuł pracy i nazwisko autora.
6. Wszystkie tablice (unikąć zbyt dużych) podobnie jak rysunki należy wykonywać na oddzielnych arkuszach i numerować kolejno liczbami arabskimi. U góry każdej tablicy podać tytuł (napis) objaśniający.
7. Po zakończeniu artykułu należy podać wykaz literatury, wymieniając w następującej kolejności: nazwisko autora i pierwsze litery imion, pełny tytuł dzieła lub artykułu, tytuł czasopisma, tom, numer zeszytu, rok i miejsce wydania oraz ewentualnie numer strony. Pozycje powinny być ponumerowane w kolejności alfabetycznej autorów; w tekście — powołania na numer pozycji w nawiasie kwadratowym, np. [3].
8. Autorowi przysługuje bezpłatnie 25 egzemplarzy odbitek pracy. Dodatkowe egzemplarze autor może zamówić w redakcji na własny koszt przy przesyłaniu korekty swej pracy.

U w a g a: Autora obowiązuje korekta autorska, którą należy zwracać w ciągu 3 dni pod adresem: Redakcja „Archiwum Elektrotechniki“, Warszawa, Koszykowa 75, Politechnika, Gmach Elektrotechniki, Zakład Elektroniki PAN. tel. 8.32.04.

Cena zł 20,—